DOCUMENT RESUME

ED 372 718 HE 027 673

TITLE	A Preliminary Study of the Feasibility and Utility for National Policy of Instructional "Good Practice" Indicators in Undergraduate Education. Contractor Report.
INSTITUTION	National Center for Higher Education Management Systems, Boulder, Colo.; Synectics for Management Decisions, Inc.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	ISBN-0-16-045151-5; NCES-94-437
PUB DATE	Aug 94
CONTRACT	RN910600-01
NOTE	68p.; For related documents, see HE 027 672, ED 363 177, ED 359 856, and ED 346 808.
AVAILABLE FROM	U.S. Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Standards; Class Size; College Environment; *College Outcomes Assessment; Curriculum; Educational Assessment; Educational Objectives; *Educational Practices; Educational Quality; Higher Education; *Instructional Effectiveness; Required Courses; Student Behavior; Student Development; Student Educational Objectives; Student Motivation; Teaching Methods; Time on Task; *Undergraduate Study
IDENTIFIERS	*Indicators; *National Education Goals 1990

ABSTRACT

This paper on the feasibility of using teaching practices as indicators of student learning was developed as part of an ongoing project to develop a process for the assessment of college student learning in the context of National Education Goal 5, Objective 5, which aims to increase the proportion of college graduates with "advanced ability to think critically, communicate effectively, and solve problems." The paper reviews the empirical research literature in three areas: (1) effects of institutional requirements including the relationship between outcomes and specific curricular requirements or coursetaking patterns, particular instructional designs, and expected levels of student performance; (2) effects of instructional practices such as class size and structure, specific classroom activities and behaviors, and influences of the institutional environment; and (3) effects of student behavior including relationships between outcomes and student time-on-task, quality of effort, and overall involvement. Available mechanisms for gathering information about educational practices and student experiences were also examined. Conclusions suggest that indicators based on student behaviors and "active learning" instructional processes gathered through student and faculty questionnaires would be most promising for development as potential national indicators, supplemented by transcript studies and assessments of typical college examinations and assignments. Contains 190 references.

HE

ED 372 718

NATIONAL CENTER FOR EDUCATION STATISTICS

A Preliminary Study of the Feasibility and Utility for National Policy of Instructional "Good Practice" Indicators in Undergraduate Education

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

U.S. Department of Education
Office of Educational Research and Improvement NCES 94-437

HE 027 673

BEST COPY AVAILABLE

A Preliminary Study of the Feasibility and Utility for National Policy of Instructional "Good Practice" Indicators in Undergraduate Education

Contractor Report

Prepared by
National Center for Higher Education Management Systems

Sal Corrallo
Project Director
National Center for Education Statistics

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Commissioner

National Center for Education Statistics

"The purpose of the Center shall be to collect, analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e–1).

August 1994

Contact:
Sal Corrallo
(202) 219–1913

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-045151-5

FOREWORD

This is the fourth report NCES has published as part of its planning to develop a process for the assessment of college student learning. The effort is in direct support of Objective 5 of National Education Goal 5. Goal 5 reads, *By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global society and exercise the rights and responsibilities of citizenship.* Objective 5 is more specific. It reads, *the proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially.*

Earlier papers reported on results of planning conferences held in November of 1991 and 1992. Participants, when asked about the direct assessment of college student learning, outlined three steps:

1. identification of the necessary set of skills and the expected levels of learning for each,

2. identification of staff training needs, and if necessary, development of instructional strategies in support of the teaching/learning of these skills, and

3. conducting direct assessment of these skills in a manner that is consistent with student learning and potential use of the identified skills.

However, several participants suggested that direct assessment of college student learning is not the only strategy available. Indirect measures, more specifically the use of good teaching practices, may also be an effective and efficient approach to assessing at least some aspects of college student learning. This paper was commissioned by NCES to review the feasibility and utility of using teaching practices as indicators of student learning. Authors Peter Ewell and Dennis Jones have written and consulted extensively on assessment in postsecondary education. They present a well researched and thorough review of good teaching practices that may be used as a measure of the curriculum students had an opportunity to learn for assessment purposes. This report is a valuable addition to the information base that will be used by NCES to develop a means of assessing college student learning.

For more information on earlier publications, or comments on this one, contact Sal Corrallo, NCES Project Planning Director, Room 306E, 555 New Jersey Avenue NW, Washington, D.C. 20208, (202) 219-1913 (Voice) or (202) 219-1801 (FAX).

Emerson J. Elliott
Commissioner of Education Statistics

ABSTRACT

A Preliminary Study of the Feasibility and Utility for National Policy of Instructional "Good Practice" Indicators in Undergraduate Education

This document addresses the feasibility of developing a range of indicators of "good practice" in undergraduate education that might be suitable for collection on a national basis consistent with the collegiate attainment objective (5.5) of the National Education Goals. Specifically, the document presents results of a focused review of the empirical research literature on general collegiate skills attainment intended to, a) assess the degree to which information on instructional practices, environments, and student experiences are reliably related to the development of critical thinking, communications, and problem-solving skills; and b) to identify and assess available data-gathering instruments and approaches that might be efficiently used to gather information about good practices in undergraduate education on a national basis.

Three specific informational domains were examined for their linkages with outcomes. The review of institutional requirements covered the relationship between outcomes and specific curricular requirements or coursetaking patterns, particular instructional designs, and expected levels of student performance. The review of instructional practices covered similar associations with class size and structure, specific classroom activities and behaviors, and influences of the wider institutional environment. The review of student behavior covered relationships between outcomes and student time-on-task, quality of effort, and overall involvement. Strongest empirical linkages were noted for "active learning" classroom practices, for broad levels of student involvement in the institutional environment, and for high student time-on-task and quality of effort. A brief review of evidence for the validity of self-reported cognitive development was undertaken that suggested moderate but reliable associations between self-reports and directly-measured cognitive skill levels. Available mechanisms for gathering information about educational practices and student experiences also were examined including administrative records, surveys of institutional practice, transcript and coursetaking methodologies, faculty surveys, and student surveys.

Overall conclusions drawn were that indicators based on student behaviors and "active learning" instructional processes gathered through student and faculty questionnaires would be most promising for development as potential national indicators, supplemented by transcript studies and assessments of typical college examinations and assignments.

PREFACE

This document provides initial guidance on the feasibility of developing statistical indicators of good practice in undergraduate postsecondary education, consistent with the need to report progress in meeting collegiate skills attainment objectives as stated in the National Education Goals. Specifically, the document examines documented empirical linkages between information on instructional practices and student experiences, and collegiate attainment; and identifies available mechanisms for gathering such information; based on a review of the research literature on student outcomes in higher education. This project is part of a larger effort by NCES to develop appropriate indicators of collegiate outcomes consistent with the National Education Goals.

The document was prepared by the National Center for Higher Education Management Systems (NCHEMS) for the National Center for Education Statistics (NCES) as Subtasks 13.2 and 13.3 pursuant to contract #RN910600.01 with Synectics for Management Decisions, Inc., under order numbers SYN-132-NCHEMS and SYN-133-NCHEMS. Section III of the document constitutes the substance of the required deliverable for Subtask 13.3, and Section IV constitutes the substance of the required deliverable for Subtask 13.2.

The principal author of the document was Peter T. Ewell, Senior Associate at NCHEMS. Contributing to the project were Cheryl D. Lovell, Paula Dressler, and Dennis P. Jones of the NCHEMS staff.

TABLE OF CONTENTS

Foreword iii

Abstract iv

Preface v

A Preliminary Study of the Feasibility and Utility for National Policy on Instruction "Good Practice" Indicators 1

I. Background and Purpose 1

II. Issues, Methods, and Sources 3

A. Defining the Outcomes Domain 4
1. Problems with Core Definitions 4
2. Interdependence of Domain Dimensions 5
3. Problems in Generalizing Across Contexts 6

B. Defining "Good Practice" 7
1. Institutional Requirements 8
2. Instructional "Good Practice" 8
3. Student Behaviors and Self-Reported Gains 8

C. Approaching the Research Literature 9
1. Sources Consulted 9
2. Limitations of the Available Research Literature 10
3. Resulting Principles for Review 11

III. Principal Findings of the Review 13

A. Institutional Requirements 13
1. General Education 13
2. Major Field 14
3. Specially Designed Courses 15
4. Levels of Expectation 15

B. Instructional "Good Practice" 16
1. Class Size and Structure 16
2. What Happens in Class 17
3. What Happens in the Wider Environment 19

C. Student Behavior 20
1. Time on Task 20
2. Involvement and "Quality of Effort" 21

D. The Special Case of Student Self-Reports 23

IV. Review of Available National Data-Gathering Instruments and Approaches 24

A. Institutional Administrative Records 25
1. The Integrated Postsecondary Education Data System (IPEDS) 25
2. Existing State Indicators System 25

B. Surveys of Institutional Practice 26
1. Academic Management Practices Survey 26
2. Seven Principles for Good Practice—Institutional Inventory 27

C. National Transcript Studies 27
1. Structure and Coherence in the Undergraduate Curriculum 28
2. The Coursework Cluster Analysis Model (CCAM) 28

D. Surveys of Faculty Teaching Practice 29
1. UCLA Faculty Survey 29
2. Seven Principles for Good Practice—Faculty Inventory 29

E. Surveys of Current and Graduating Students 30
1. Learning and Study Strategies Inventory (LASSI) 30
2. Motivated Strategies for Learning Questionnaire (MSLQ) 31
3. Cooperative Institutional Research Program (CIRP) Surveys 31
4. College Student Experience Questionnaire (CSEQ) 31
5. ACT-ESS Alumni Questionnaire 32
6. NCHEMS SOIS Questionnaire 32

V. Summary and Recommended Actions 33

Summary Chart of Potential "Good Practice" Indicators 36

References Cited in the Text 39

References Consulted But Not Cited in the Text 52

A Preliminary Study of the Feasibility and Utility for National Policy of Instructional "Good Practice" Indicators in Undergraduate Education

The National Center for Higher Education Management Systems (NCHEMS)

I. Background and Purpose

Since 1990, resource groups centered on each of the National Education Goals have met repeatedly to help determine how progress in attaining them might systematically be tracked on a national basis, and a number of preliminary progress reports have been issued. Among the most challenging goal areas in this respect has been collegiate achievement--in the language of the Goal 5.5, the ability of graduating college seniors to "think critically, communicate effectively, and solve problems."

Concerns about collegiate achievement, especially in its general education component, have also been raised in other quarters. Since the mid-1980's, pioneered by the *Involvement in Learning* report of the Department of Education's Study Group on the Conditions of Excellence in American Higher Education (NIE 1984), both institutions and state governments have undertaken initiatives to assess and improve the common component of the college curriculum. Partly this is response to internal academic concerns, but more insistently it is a reaction to the increasing salience of communications and problem-solving skills in future workforce needs--needs as expressed, for instance, in the report of the Secretary's Commission on Acquiring Necessary Skills (SCANS) of the Department of Labor. References to this wider set of issues suggests that an appropriate national assessment of collegiate skills-- regardless of what is explicitly called for in the National Goals--cannot proceed in isolation.

To date, the primary thrust of national discussion regarding the assessment of collegiate skills has centered on the development of a performance-based examination. It is the primary recommendation of the Goal 5 Resource Group that such an examination be developed and in-depth conferences on its content and design involving background papers and extensive discussion have already taken place under the sponsorship of the National Center for Education Statistics (NCES). These activities confirm that the development of such an examination will be a long and difficult process. Among the many issues involved in its development will be the following:

- high costs and long timelines for development. Past experience in developing meaningful postsecondary assessments of collegiate skills of the kind recommended by the National Goals Panel suggests that a timeline of at least five years will be required. This experience also suggests that to develop, validate, and pilot the kinds of complex, performance-based exercises required for such an examination will be an expensive undertaking. The technical properties of such assessments are complex and

in many cases unknown: standard validity and reliability measures are hard to apply and the results are often subject to unknown and uncontrollable biases. Growing experience with assessing college student populations, moreover, suggests that motivating students to participate and to do their best may be a major problem. All these challenges, no doubt, can be overcome, but the process will take time and money. Meanwhile, if information about higher education is to inform national policy, it must be collected by other means.

- the need to gain consensus about specific domain abilities. As noted by the Goal Five Technical Panel, the first step in designing an appropriate national assessment of collegiate skills will be a consensus-building process. Identified abilities of "critical thinking, communication, and problem-solving", while broadly understood, are operationally ill-specified and are currently embodied in quite different kinds of assessments. Building consensus will take time: a recently-issued NCES RFP intended to begin the assessment design process by specifying domain, for instance, allocates eighteen months to this process. Wider conversations about future workforce needs and the need to improve practice in undergraduate study more generally, moreover, suggest that the specific domain content of the National Goals should not be given too narrow a construction. In the light of these larger public issues, discrete tests of disembodied abilities--no matter how technically sound--cannot in themselves meet the needs of future national policy.

- the need to direct policy action. One historical drawback to the utility of cognitive assessment results in the improvement of postsecondary education-- despite their often sound construction--is the difficulty of linking such results to prior educational experiences that can be effectively manipulated through policy action. Valid test scores can indicate rather precisely what has been accomplished and where deficiencies exist, but they often provide little guidance about what can and should be done. This appears particularly to be the case for the kinds of broad, higher-order abilities noted in the National Goals.

These drawbacks have led a number of observers--including the original National Goals Panel Resource Group for Goal Five--to recommend development of additional national indicators of instructional practices in higher education to supplement more direct cognitive assessments of collegiate abilities. This case rests on two main grounds. First, such indicators may provide important additional information which can help policymakers make sense of the findings of end-point assessments. If their validity can be established, such indicators might not only "supplement" information derived from direct performance measures, but can be useful in their own right. Except as a pure benchmark of progress, it makes little policy sense to collect outcomes information in the absence of information on key processes that are presumed to contribute to the result. Higher education, moreover, is particularly in need of national information about contexts and processes because of its enormous variety. Not only do colleges and universities exist in many forms with diverse

educational missions, but unlike the situation typical of K-12 education, there is little curricular commonality in the experiences of college students. Given this situation, information on outcomes alone is virtually uninterpretable in the absence of information about key experiences.

Second, a set of indicators tied to important instructional practices provides clear policy leverage for action. A major difficulty of outcomes-based performance funding experiments in higher education has been has been relative lack of direction about where to invest resources to obtain best results (see for example, Banta 1986). More effective in causing meaningful change have been categorical or marginal funding mechanisms targeted directly on fostering instructional actions that previous research has found to be effective (Jones and Folger 1993). A set of national benchmarks about key experiences and conditions in general education, therefore, might be of considerable value in determining the degree to which colleges and universities are willing and able to act consistently with the national goals.

The purpose of this document is to present results of a background study intended to assess the initial feasibility of collecting additional benchmark data of this kind. Specific purposes of the project were:

- to explore through a review of the research literature on college outcomes, the validity of "good practice" indicators as proxy measures of student cognitive attainment and/or as important sources of policy information to help achieve higher levels of cognitive attainment.

- to identify specifically a number of potential data-gathering mechanisms that incorporate "good practice" indicators, and present their strengths and weaknesses as candidates for eventual inclusion in a national data- collection system to track progress on Goal 5.5.

Results of these tasks are presented in Sections III and IV respectively, which constitute the core of the document. Section II provides an overview of sources consulted and the methodological issues that arose. Section V summarizes the study's conclusions and reviews some next steps that should be considered.

II. Issues, Methods and Sources

A study of this kind by its very nature will encounter at least three obstacles. First, the specific contents of the cognitive "outcome" domains in question--"critical thinking, communication, and problem-solving"--have not yet been fully identified. Defining this domain more precisely is a principal task of the recent NCES RFP, and although much preliminary work has been done in the context of two NCES study-design workshops, much remains to be accomplished. Meanwhile, considerable empirical work has already proceeded

in the absence of a commonly-defined criterion variable. Second, the range of possible "good practices" and conditions that might be explored in conjunction with this domain is vast. Choices, therefore, had to be made about which particular areas appeared most promising from both an empirical and a policy point of view. Finally, the literature on collegiate outcomes is large, diffuse, and methodologically uneven. Fortunately, it has also been reviewed before on a number of occasions--though not with such a specific purpose in mind. As a result, this review was able to benefit greatly from prior work in determining particular empirical studies to examine in greater depth.

A. Defining the Outcomes Domain

Despite considerable discussion, disagreement remains about the precise specification of the abilities noted in Goal 5.5. In brief, these issues are of three kinds.

1. Problems with Core Definitions. Multiple paradigms for specifying the content of "critical thinking and problem-solving" have arisen historically, owing largely to the different disciplinary roots of particular approaches. Recently, two NCES study-design workshops made considerable progress in achieving consensus on the nature of this underlying ability, including the recognition that "critical thinking" and "problem-solving" should be viewed in the context of a single, comprehensive construct (Facione 1992).

Absent this recent consensus, differences in perspective have, of course, colored past empirical work on the topic. There are a number of useful reviews of critical-thinking definitions that outline these distinctions (e.g., Jones 1993a, McMillan 1987, Kurfiss 1988). Kurfiss' (1988) analysis, for example, reveals three main topic areas--a) "informal logic" that principally involves the construction and critique of arguments, b) "cognitive processing" that principally involves how an individual "constructs meaning" from a given set of information, and c) "development of intellect" that principally involves identifiable changes in a given individual's metacognitive attributes through a number of distinct but increasingly sophisticated "stages" over time. The majority of such conceptualizations of critical thinking, however, have evolved formally and individually-- that is, their specification resulted from an a priori set of individual philosophical positions, each resulting in a unique taxonomy of abilities. An alternative approach to defining the domain is empirical. A project recently commissioned by the American Philosophical Association (APA), for instance, evolved a consensual description of critical thinking as "purposeful, self-regulatory judgment" involving the possession and deployment of both cognitive abilities and affective dispositions, using a multi-round Delphi process conducted among forty-six critical-thinking scholars (Facione 1990, 1992). Among the specific cognitive abilities noted in this formulation are, interpretation, analysis, evaluation, inference, explanation, and self-regulation. Basic contents of the domain identified in this project were broadly confirmed by the NCES Study Design Workshops.

Whatever their particular views on details of the ability, the majority of those attempting to delineate the domain of critical thinking have maintained that its underlying properties are both general and teachable: that is, they can be applied to a wide range of problem-solving situations and they can be positively influenced by instruction designed specifically to enhance these abilities (e.g., Paul 1992; Paul, Nosich and Fisher 1992; Halpern 1992; Facione 1992). Most also note that true possession of the ability consists of both a basic competence and a disposition to use it appropriately (e.g., Facione 1990, Gray 1993). Paul's (1984) notion of "weak-sense" critical thinking, for instance, rests largely upon a construction of the ability as a "technical" skill that can be deployed or not at will, while his view of critical thinking in its "strong sense" involves an imperative to act appropriately (including the disposition to consciously adopt and examine the points of view of others) that cannot be so avoided. Ennis (1985) also argues that the test of action is key--and that critical thinking is visible primarily in "decisions about what to believe or do." They agree, moreover, that its primary area of deployment is broad and non-specific--in addressing what Paul (1984) terms "the messy problems of everyday life" (p.5).

Definitional complexities of this kind are far fewer for the domain of communications. Most commentators, however, recognize several distinct dimensions of this ability, including being able to receive information as well as simply transmit it (e.g., reading and listening skills) and a distinction between formal and group-oriented (interpersonal) communication (Jones 1993b). As in the case of critical thinking, most also add judgmental and dispositional qualities to purely technical capacities by including such attributes as a sense of audience or the ability to shift communications strategies to match the needs of changing contexts (Daly 1992).

2. Interdependence of Domain Dimensions. The other side of attaining broad definitional consensus, of course, is the common inability to distinguish effectively among various domain dimensions. This difficulty is the most pronounced for critical thinking and problem-solving--although there is a slight tendency for commentators to view the latter as slightly more "applied" and quantitative (Jones 1993a; Kurfiss 1988). Viewed operationally, however, many believe the underlying qualities of these two Goal 5.5 traits to be essentially the same (e.g., Nummedal 1991, Halpern 1992), a conclusion also reached by the empirically-based Delphi study, conducted under the auspices of the APA (Facione 1990). Once communication moves beyond the technical ability to transmit information, moreover, it also becomes strongly intertwined with analytical skills (Daly 1992). Good communicators, most agree, require the ability to synthesize material, to draw appropriate conclusions, and to analyze the messages of others (Jones 1993b). "Explanation," moreover, is cited as one of the six core critical thinking skills in the Delphi Study (Facione 1990). Indeed, a conclusion reached by the two NCES study-design workshops was the need to conceptualize Goal 5.5 abilities from a unitary perspective--as a set of related sub-abilities within a single core concept.

Many of these interdependencies are empirically visible in strong empirical correlations between performances on "critical thinking" tasks and more general measures of ability

(McMillan 1987). McPeck (1981), for instance, shows strong associations between performance on the commonly-used Watson-Glaser Test of Critical Analysis and more general IQ measures. Recent studies that cross-tested students on a range of commercially-available collegiate general education assessment instruments (e.g., Thorndike 1990, Banta and Pike 1989) have found much the same thing: the results of each are highly related to the rest no matter which "subskill" is being purportedly measured, and the correlations of all with assessments of prior ability are substantial. Similar relationships have been established between performance on a range of critical thinking instruments and a student's stage of cognitive development as operationalized through the Bloom developmental taxonomy (Cano 1993, Cano and Martinez 1991), or through the Perry-based Reflective Judgement Interview (Mines, King, Hood and Wood 1990). Such interdependencies, together with the extremely broad nature of the ability being specified (Pace 1979), probably account for the relatively small gains in critical thinking that have been reported empirically among college students (e.g., McMillan 1987; Whitla 1978; Winter, McClelland and Stewart 1981; Pascarella 1989) after statistical controls for prior student ability are applied.

3. Problems in Generalizing Across Contexts. A final area of some dispute is the degree to which learned "generic" abilities can in fact be deployed effectively across different settings or whether, in fact, altered contexts define a new and different skill. Critical thinking commentators tend to conceptually duck the question by including the ability to shift contexts as an explicit attribute of the trait itself (e.g., Paul 1984, Ennis 1985). Empirically, however, there is evidence that this is not only difficult, but perhaps inappropriate as well. Recognizing this difficulty, more recent conceptualizations of critical thinking have incorporated the need for domain-specific knowledge and how it should be appropriately deployed as a central aspect of the ability (e.g., Facione 1990).

Probably the most commonly-cited problem is that of generalizing such abilities across disciplines (Kurfiss 1988, McMillan 1987). Considerable experimental evidence suggests that the actual operations and constructs labelled "critical thinking" in each discipline may in fact be distinctive, and ought therefore to be assessed independently (e.g., Campione and Armbruster 1985; Weinstein 1990; Chipman, Segal and Glaser 1985). Illustrating the difficulty nicely is a study by DeLisi and Staudt (1980) in which students in different disciplines attained equivalent levels of "formal reasoning" but only by doing well in problems related to their own fields of study. As a result, many "general" measures of cross-cutting collegiate abilities attempt deliberately to spiral items and tasks from a wide range of disciplinary contexts in any given assessment.

Other elements of context have also been raised as having a marked affect on the assessment of generic critical-thinking/problem-solving skills. Reflecting the notion of "disposition" as an important component of the core ability, for instance, students may choose not to use skills previously learned in one context simply because they are not actively prompted to do so in another (Perfetto, Bransford and Franks 1983). In short, "contextual validity" remains a problem for the assessment of any generic ability, and generalizations about what is

empirically linked to this ability should be made with special care (Mentkowski and Rogers 1988).

Given these issues, and the fact that common definitions of the abilities noted in Goal 5.5 have not yet emerged, the course chosen for this focused review was highly inclusive. Any study that attempted to link particular attributes of the instructional environment or behavior with broad thinking or analytical skills using appropriate statistical controls was considered a candidate for inclusion.

B. Defining "Good Practice"

A similar problem was encountered in delimiting factors potentially related to core abilities that might usefully serve a national indicators system. While cognitive indicators are conceptually straightforward (at least at first), many different kinds of statistics have been proposed as "indirect" measures of academic progress.

A first important point, of course, is that all indicators of educational attainment are in some sense "indirect." Purpose-built, performance-based assessments of areas of knowledge and skill are no exception, and must not be confused with the actual entity that they purport to represent. But less immediate varieties of potential indicators are of two very different kinds (Ewell and Jones 1991). On the one hand, useful and reliable statistics for "tracking progress" might be developed that are highly correlated with cognitive abilities, without being causally related to the enhancement of these abilities. Examples include the results of other, more easily administered, examinations known to be related to the traits in question, and student self-reports. "Proxy" indicators of this kind are commonly used in other fields to document trends, but their major drawback is that they cannot generally be used to inform policy. On the other hand, useful indicators might be developed consistent with an underlying philosophy that knowledge of educational inputs and processes is worth obtaining in its own right, so long as evidence exists that these are broadly related to cognitive attainment. If such an approach is followed, the policy dividend is not only the opportunity to track progress but also to inform action. In identifying potentially useful indicators of "good practice," the latter is the predominant focus of this paper, but the former is far from excluded.

Specific classes of potential indicators consistent with this philosophy were identified through a number of sources. These included national reports on the hallmarks of effective undergraduate instruction (e.g., NIE 1984, AAC 1985, Chickering and Gamson 1987), as well as widely influential research-based syntheses of practices designed to improve collegiate teaching and learning (e.g., Astin 1985, Gamson and Associates 1984). Together, these sources suggested three broad categories of potential interest in the development of "good practice" indicators (Ewell and Jones 1991):

1. Institutional Requirements. Indicators in this class would be intended to address the degree to which undergraduate requirements contain curricular features expected to be associated with collegiate attainment in the areas of critical thinking, communications or problem-solving. Examples include:

- specific proficiencies required for attainment of the baccalaureate degree--for example, explicit demonstrations of writing, speaking, computational ability, foreign language proficiency, etc.

- specific types of experiences required for attainment of the baccalaureate degree--for example, is it possible for students to graduate without having written a major research paper, taken a math course, taken a laboratory science or taken a foreign language?

- specific "capstone" or other integrative experiences required for graduation--for example, an internship, problem-oriented senior seminar, or senior thesis or project.

2. Instructional "Good Practices." Indicators in this class would be intended to address the degree to which typical student instructional experiences are consistent with established principles of good practice in undergraduate teaching (e.g., Chickering and Gamson 1987, Angelo and Cross 1993)--for example "active learning", frequent "feedback" on performance, or frequent student/faculty contact. Examples include:

- typical class-sizes encountered in lower-division courses--for example, how likely is it that a lower-division student (or first- term freshman) is enrolled in at least one class with fifteen or fewer students, in which "active participation" is likely?

- instructional experiences reported by students as typical of their undergraduate coursework--for example, frequency of writing or speaking required, levels of participation in group study or explicit problem-solving experiences, amount and type of out-of- class work required per week, numbers of assignments requiring outside independent work, or the proportion of course final examinations taken that required an essay or problem-solving component.

- additional out-of-class experiences reported by students as typical of their undergraduate experience--for example, frequency of out- class-contact with faculty members, active participation in faculty research projects, participation in on- or off-campus work related to their course of study, participation in group study, frequency of independent college-related research or study, or frequency of tutoring another student.

3. Student Behaviors and Self-Reported Gains. Indicators in this class would be intended to address the degree to which students report behaviors and outcomes consistent with good

practice in undergraduate instruction--and particularly in the acquisition of critical thinking, communications, or problem-solving skills. Examples include:

- student use of time ("time on task") in selected areas--for example, reading, writing, working mathematical or scientific problems, talking in class, talking with other students about class-related material, or working on independent research or library assignments.

- student self-reported gains in selected areas--for example analytical/problem-solving skills, oral and written communications, ability to think critically, or ability to work cooperatively.

- student self-reports regarding their reactions to college-level work--for example, the proportion of current college students reporting being actively challenged by their classes or out-of-class assignments, or the level of self-reported interest and involvement in academic matters.

This initial taxonomy of potential "good practice" indicators was used explicitly to guide the review process reported in Sections III and IV below.

C. Approaching the Research Literature

The research literature on collegiate outcomes is both large and varied, and it has been reviewed many times before. As a result, a two-staged strategy was employed to conduct this focused review--beginning with more general treatments of the literature to identify a smaller set of promising studies and evidence-gathering approaches for investigation in greater depth. Available empirical literature is also affected by a number of typical methodological caveats that limit the conclusions that can be drawn. As a result, several principles were developed for conducting the review and for determining whether and how particular findings should be reported.

1. Sources Consulted. As noted, a two-staged process was used to identify individual studies that might establish empirical linkages between one or more identified "good practice" elements and particular cognitive outcomes consistent with Goal 5.5. A similar process was used to identify and inventory individual instruments or methods that might be used to gather data as part of a national indicators system.

The first stage of the search process involved examining extant reviews of empirical studies. Most helpful here were recent comprehensive reviews of the collegiate outcomes literature by Pascarella and Terenzini (1991) and Pascarella (1985). Second in importance were a set of somewhat older compendiums of studies of collegiate impact including those of Feldman and Newcomb (1969), Bowen (1977), Pace (1979), and Lenning and Associates (1974). While broadly useful, findings reported here were generally limited by the fact that relevant studies

were conducted more than twenty years ago and were largely confined to traditional, residential, full-time student populations. A final group of secondary sources consulted at this stage consisted of specialized reviews of findings within specific subject areas like critical thinking (e.g., McMillan 1987; Beck, Bennett, McLeod, and Molyneaux 1992) or communications (e.g., Daly 1992), that summarized results from the point of view of effective teaching practice (McKeachie, Pintrich, Lin and Smith 1986; Cole 1982), or that reviewed findings from a particular policy perspective (e.g., Ewell 1988).

In parallel, computer searches were run on the ERIC Clearinghouse database and other periodicals indices, using keywords such as "critical thinking," "analysis," "thinking skills", "problem-solving," "communications," and "cognitive development." These uncovered additional unpublished studies of interest, as well as others that were sufficiently recent that they could not be included in published reviews. At the same time, tables of contents for the last two years were examined for each of the higher education research journals, and for leading journals in educational research and educational psychology.

Specific studies that investigated empirical linkages between one or more "good practice" factors and cognitive outcomes arising from this preliminary review were then examined in greater depth to determine, a) the nature and strength of the association (e.g., direct or indirect), b) the instruments used and the study's setting (including both student population and institutional context), and c) particular aspects of the study design employed (e.g., the use of before/after measures, experimental groups, or particular control variables). Findings were then organized in terms of the taxonomy of "good practice" factors presented earlier for inclusion in Section III. References cited and consulted through this process are provided at the conclusion of the document.

2. Limitations of the Available Research Literature. Although large and diverse, the body of empirical work on collegiate outcomes is methodologically limited in many ways (Pascarella and Terenzini 1991, Pascarella 1991, Terenzini and Pascarella 1991, Pascarella 1985). First, relatively few studies exist that are multi-institutional, longitudinal, and contain a full array of control variables. This means that generalizations across studies must be made with great care, because contexts may vary and observed effects may be due to factors not explicitly investigated. Especially troublesome is the common absence of appropriate comparison groups to investigate causal relationships: studies that employ non-college-attending populations to control for the influence of maturation (e.g., Pascarella 1989) are extremely rare. The effects of particular experiences or settings, moreover, may be indirect or may be mediated by other variables rather than being directly observable (Pascarella 1991, Pascarella and Terenzini 1991). And in many cases, causal direction per se may be impossible to determine: do students, for instance, perform better in college because they become more involved or are they induced to become more involved because they are doing better (Pike 1991)?

When the outcomes in question are cognitive (rather than affective or behavioral) additional issues arise. Most troublesome here is the fact that the net effects of college attendance on

generic cognitive attainment appear modest after prior ability is controlled. Pascarella and Terenzini's (1991) exhaustive review of the most recent twenty years of college impact research, for instance, broadly confirms Bowen's (1977) earlier estimate of raw effect sizes on verbal skills of about half a standard deviation after college attendance-- approximately half the parallel gain estimated for subject-area knowledge. They attribute much of this difference to the ambiguities associated with measuring broad traits of this kind, but report that even these modest gains tend to diminish after pre-college characteristics are controlled for.

When the outcomes domain is explicitly confined to "critical thinking", moreover, reported gains are even more limited, and in many cases disappear altogether (McMillan 1987; Beck, Bennett, McLeod, and Molyneaux 1992). Attempting to "parcel out" these relatively small net effects among a number of discrete elements of behavior and environment will often yield little of significance, even if broad associations among these factors are in fact present. As Pace (1985) puts it, "if you choose to think big about the scope and significance of outcomes, then you must also think big about the magnitude of college experiences when you seek explanations of outcomes (p.17)." As a result, many of the studies reviewed claimed general effects consistent with a cluster of characteristics and experiences (e.g., Astin's "theory of involvement" [1985, 1993, Pascarella 1989]; Pace's concept of "quality of student effort" [1984, 1990]; or Winter, McClelland and Stewart's "Ivy Experience" [1981]), but were often unable to document unique empirical linkages for individual characteristics or behavioral elements.

3. Resulting Principles for Review. Definitional uncertainties about cognitive domains and methodological limitations of the existing research literature suggested some principles for conducting the review. Briefly, these were as follows:

- broad focus. Potential interdependencies among the specific domain elements of critical thinking, communication and problem-solving, together with demonstrated relationships between these qualities and more general cognitive abilities first imply that the "dependent variable" be constructed as broadly as possible. This demands examining general higher-order cognitive skills that are labeled in many different ways--including "thinking" and "analytic" skills, "metacognitive" skills, and the development of general verbal and quantitative abilities. Generally excluded, however, should be studies examining knowledge growth in specific discipline content or overall collegiate achievement (as reflected, for instance, in grade performance). This principle also allows the inclusion of studies that operationalize or measure these skills in different ways--including classic "critical thinking" assessment instruments like the Watson-Glaser Critical Thinking Appraisal or the Cornell Test of Critical Thinking; more general college-level examinations like the Graduate Record Examination (GRE), the ACT Assessment, and the National Teacher's Examination (NTE); or more recently-developed "general education" assessment instruments like the ACT-COMP, ACT-CAAP, the ETS Academic Profile, and the College- BASE examination. The results of assessment using more open-ended methods such as

developmental interviews (e.g., Knefelkamp 1974), essays or other performance measures (e.g., Whitla 1978; Mentkowski and Strait 1983; Winter, McClelland and Stewart 1981), and self- reports of student progress were also considered (e.g., Pace 1984, 1990; Astin 1977, 1993). In short, the strategy was to cast as wide an analytical net as possible, consistent with the core concepts.

- robust effect. The decision to choose a broad focus implied in turn the need to report primarily on those empirical linkages that appeared to persist across quite different settings, and that seemed verified regardless of how core concepts were operationally defined and measured. While in few cases was this true, the principle of "triangulation" nevertheless appeared justified because of substantial variations in what outcomes were being measured, in what contexts, and how. This principle also argued that overall weight of evidence should count for at least as much as classic methodological rigor in identifying promising alternatives. That is, if a given linkage emerged consistently across a number of single-institution studies or studies containing methodological limitations, or that showed consistent effects in a particular direction without achieving statistical significance, this was worth noting.

- establishing linkage, not just cause. These relaxed methodological constraints are justified by the review's primary intent. To be useful as an indicator, a given statistic need not be directly or causally related to the entity of interest; the minimal requirement is that it reliably covary with this entity. Most of the well- recognized methodological limitations of the existing literature (e.g., Pascarella and Terenzini 1991, Terenzini and Pascarella 1991, Pascarella 1985) center on the absence of particular study- design elements needed to establish effect priority, not the presence or magnitude of a particular association. To be sure, clearly establishing patterns of cause and effect is extremely useful for policy purposes where this can in fact be done. But equally valuable for charting progress (and quite present in the research literature) may be any "hallmarks" of a sound undergraduate experience that the evidence suggests.

- implied utility. This last observation helps frame a final principle--that useful indicators should be linked not only to the underlying phenomena that they are designed to reflect, but also to policy action and understanding. To pass this test, any proposed indicator must meet a number of conditions beyond those that can be verified empirically. First, it must reflect broadly representative practices or conditions. Although student time on task and participation in a specific set of programmed instructional methods (PI or the so-called "Keller Plan", for example) both appear to have demonstrable causal relationships with cognitive growth in college (Pascarella and Terenzini 1991), the former is much more broadly useful as a national indicator. Second, a useful indicator must reflect processes or phenomena which are potentially important in their own right. Because of its association with additional items of public interest like costs and general "consumer satisfaction", for instance, an indicator based on factors like student-faculty interaction and the accessibility of

faculty may be of considerably more utility than simply the fact that these factors are linked empirically with cognitive growth in college. Finally, a potentially useful indicator must be able in some fashion to help inform policy action--either by enabling broader public understanding of a given condition or by suggesting a mechanism for change. Absent this criterion, for instance, the best single indirect measure of the three Goal 5.5 outcomes suggested by the research literature is attainment on virtually any assessment of incoming student ability.

Applied together, these four principles provided a reasonable guide for reviewing a diverse literature. At the same time, as described in Section III below, they were strongly discriminatory with respect to the initial taxonomy of "good practice" indicators, and resulted in a relatively small body of potentially useful measures.

III. Principal Findings of the Review

Results of the literature review are summarized in this section according to the taxonomy of potential "good practice" indicators presented earlier. In some cases, however, the taxonomy did not adequately reflect factors that emerged in the research literature as strong empirical correlates of improved learning, and additional subsections were added to reflect these factors. For the most part, these consisted of more general, cross-cutting aspects of the educational setting or experience. This section broadly summarizes the review itself, and the recommended direction with regard to indicators development that it suggests is provided in Section V.

A. Institutional Requirements

Very little of the research literature directly addresses linkages between general institutional requirements and the attainment of critical thinking or communications skills (Pascarella and Terenzini 1991). Investigations that have been related to this general topic include, a) assessments of the effects of different kinds of general education requirements or coursetaking patterns, b) assessments of the differential effects of major curricula, c) assessments of the impact of particular course/curriculum designs, and d) investigations of the relationship between typical levels of difficulty or proficiency expected and general cognitive growth.

1. General Education. "General Education" is a sufficiently broad term that it has resisted operationalization in most empirical studies. There is, however, modest evidence that overall exposure to a wide range of academic material is related to higher levels of attainment on general measures of collegiate ability. Using the ACT-COMP examination that involves the deployment of high-level analytic abilities, Forrest and Steele (1982) found in a study of 44 institutions that gains in performance were related to the overall "breadth" of undergraduate general education requirements. Other studies using the same instrument, however, have failed to replicate this finding and have cited methodological problems with its use of

estimated gain (Banta, Lambert, Pike, Schmidhammer and Schneider 1987). Earlier studies (generally restricted to a traditional residential student body) also documented growth in critical thinking, as assessed through the Watson-Glaser, with the curricular features of a classic "core" curriculum (e.g., Dressel and Mayhew 1954, Gaff 1983); "curriculum flexibility" has also been noted as a factor associated with general cognitive growth by Centra and Rock, using GRE residual scores (1971). Additional studies have documented associations between self-reported cognitive growth and such curricular factors as free choice of electives or participation in an honors program (e.g., Astin 1977, Pace 1979). A more sophisticated replication of these findings is reported by Astin (1993) in a longitudinal design involving self-reports, standardized examinations, and faculty surveys. Among the factors noted as especially related to growth in critical thinking were participation in interdisciplinary and honors coursework--though both were included within the rubric of a wider "humanities orientation" for the institution. It should be remembered, however, that all these reported gains for general ability are modest compared to growth in subject-area knowledge and that the evidence for the effectiveness of any given curricular structure is fragmentary.

Studies have also begun to examine the differential effects of specific behavioral patterns of coursetaking. Zemsky (1989) has developed a methodology for examining the "breadth" and "depth" of undergraduate coursetaking experiences using national samples of transcripts, but the resulting metrics have not yet been linked to learning. The "differential coursework methodology" employed by Ratcliff and Associates (1988), however, has detected significant differences in performance on different types of GRE items that can be traced to the types of courses that students have taken; effects were detected here for both specific disciplines and for the level at which a given course was taken. Coursework coverage also appears to have an effect on subscale performance on the ACT-COMP (Pike and Phillipi 1988, Pike 1989, Pike and Banta 1989), and subscale attainment on such instruments as the Watson-Glaser (Annis and Annis 1979). The enormous variety of coursetaking patterns detected by Ratcliff and Associates (1988) on the one hand helps to explain why so few "general" curricular effects tend to emerge from the empirical literature; the differential pattern detected, on the other hand suggests that the kinds of courses students enroll for do make a difference.

2. Major Field. Evidence of differences in the outcomes associated with majoring in different fields have been widely reported over forty years of research (Pascarella and Terenzini 1991, Bowen 1977, Pace 1979, Feldman and Newcomb 1969). For the most part, however, these have shown variations in individual subscore or subskill performance that are particularly associated with a given field of study, rather than demonstrating an overall advantage for particular majors in fostering in generic abilities. DeLisi and Staudt (1980) for instance obtained comparable total scores on formal reasoning ability for students across disciplines, but students did better in problem contexts related to their own majors. Whitla (1978) obtained similar results on the performance-based Test of Analysis and Argument and Test of Thematic Analysis for matched samples of freshmen and seniors at three institutions: students learned better and faster in areas related to their primary fields of study. Rather than supporting arguments about the efficacy of particular curricular experiences or

structures, however these effects seem more likely to be the result of different levels of exposure and sheer time on task in particular areas.

3. Specially-Designed Courses. A large number of studies have demonstrated the impact of special types of instructional design on knowledge gain. These include variations of self-paced, computerized or programmed learning experiences, or individualized instructional techniques (Pascarella and Terenzini 1991). Where courses in question are specifically designed to accomplish a given score gain, they will often have that effect. For instance, Chaffee (1992) reports gains pre-post in critical thinking ability after students were exposed to specially-designed critical thinking courses "paired" with subject area courses. But viewed overall, results are mixed. In McMillan's (1987) review of seven such studies, only three demonstrated significant associations between courses and outcomes, and these were modest. Beck, Bennett, McLeod and Molyneaux (1992) reviewed ten similar studies in the restricted area of nursing education and found no significant associations between curriculum structure and critical thinking gains.

Examining this body of work more closely, it appears that any documented effects of this kind may be as much due to what is done in the classroom with respect to active learning and the provision of frequent feedback than to any particular elements of course design. In a typical study, Widick, Knefelkamp, and Parker (1975) for example demonstrated developmental gain on the Perry scale after students participated in a class designed around such activities as debates, role-playing, and the use of learning logs. Both because of this possible interdependence and because they will be more generally applicable therefore, indicators based on in-class activities themselves appear more promising than those that merely report the presence of specially-designed classes or curricula.

4. Levels of Expectation. In both K-12 and postsecondary education frequent claims have been made that heightened levels of expectation for all students will induce better performance (Wiggins 1989, NIE 1984, Chickering and Gamson 1987). But few systematic empirical investigations of this proposition have been undertaken. At the institutional level of analysis, quite a number of studies have demonstrated a linkage between gains in critical thinking and attendance at highly-selective institutions--especially small private liberal arts colleges (e.g., Whitla 1978; Winter, McClelland and Stewart 1981). A similar pattern is also reported for self-reported gains on these skills (e.g., Astin 1993; Pace 1990, 1984). Because a number of these studies have used students at non-selective colleges as control groups, it is possible to claim greater enhancements of higher-order skills for those attending selective institutions. But whether this is due to the nature of the students or the experiences that they encounter cannot be directly determined (Pascarella and Terenzini 1991).

There is certainly strong evidence that selective colleges do provide challenging environments, however, and methods are available for assessing the degree of challenge. Braxton and Nordvall (1985), for instance, used the Bloom taxonomy to classify the difficulty of typical examination questions administered at 52 small private colleges and found considerable differences in cognitive level based on institutional selectivity. Pintrich

(1988) proposes a similar method for coding the difficulty of posed examination questions, though it has not yet been used in outcomes studies. Finally, Fischer and Grant (1983) established differences in the cognitive level of student-faculty exchanges across classrooms at several institutions, but these differences may have been the result of selectivity. Taken together, these findings suggest that national samples of what students are typically asked to accomplish might provide useful information to track progress regardless of whether or not high expectations "cause" greater learning (e.g., Cheyney 1991).

In general, then, the empirical literature provides only mixed evidence of a relationship between higher-order skills and structural or curricular aspects of instruction. Most of the effects that have been demonstrated for curricular structures--with the possible exception of level of challenge--may equally be a result of what happens differently in the classroom within such curricula or what students actually do as a result. And, as reported below, these are both areas for which there is substantial independent evidence of impact.

B. Instructional "Good Practice"

In contrast to the case for institutional requirements, a substantial literature documents empirical linkages between particular elements of instructional delivery and improvements in learning (Pascarella and Terenzini 1991; McKeachie, Pintrich, Lin, and Smith 1986). Lists of effective practices that overlap substantially with most of these elements have also been advanced from widely different sources, ranging from educational researchers (e.g., Astin 1985), representatives of the critical thinking movement (e.g., Paul 1992), and national blue-ribbon bodies (e.g., NIE 1984, AAC 1985). Specific bodies of empirical work around this general topic can be grouped into a number of clusters, including, a) class size and structure, b) what happens in class, and c) what happens in the wider institutional environment.

1. Class Size and Structure. Despite public perceptions, the evidence that class size or student/faculty ratio has a direct impact on learning in college classrooms is slim. An extensive review of findings from K-12 classrooms undertaken in 1988 does not support a general linkage between class size and cognitive gain (OERI 1988), and probably the largest single postsecondary study of this topic in a college setting corroborates this finding (Williams, Cook, Quinn and Jensen 1985). The same in general can be said for broadly-described elements of class structure--for instance the familiar distinction among lecture, discussion, and laboratory modes of delivery (McKeachie, Pintrich, Lin and Smith 1986) at least with respect to learning content--although alarming rates of atrophy in learning have also been reported for the lecture method (McLeish 1968). Balancing this general finding to some extent is evidence of the effectiveness of smaller classes in developing higher-order abilities--particularly in communications and critical thinking (e.g., Schalock 1976, McKeachie 1980). An overwhelming finding, moreover, is that students are more satisfied with smaller classes and report extensively that they learn more. Astin (1993), for instance, documents only very indirect empirical linkages between overall student/faculty ratios and actually-assessed student learning outcomes, but strong direct effects on students'

overall satisfaction with the college experience. Similarly, a statewide study in Florida presented consistent student claims to have learned more in smaller classes across a wide range of settings and subjects (State of Florida, Postsecondary Education Planning Commission 1990). Interpreting these and other results, Pascarella and Terenzini (1991) advance the hypothesis that these reported linkages are present, but are actually the result of the opportunities for active learning, frequent feedback, and for the practice of learned skills that smaller classes and discussion sections may afford; simply offering such instructional modes may be insufficient.

2. What Happens in Class. The empirical literature provides broad confirmation that general cognitive growth is associated with specific types of classroom activities and instructor behaviors--far more so than for curricular structures or requirements. In many ways, given the extreme variations in how students "act out" designed curricula through actual coursetaking (Ratcliff and Associates 1988) and likely variations across otherwise similar classrooms in instructional approaches and styles, this is not a surprise. Indeed, a number of reviewers of this literature have emphasized that real effects on student learning tend only to emerge in identifiable "micro-settings" where they are less likely to be masked by sheer contextual variability (e.g., Pascarella and Terenzini 1991, Ewell 1988, Pace 1985).

A major body of findings under this heading is consistent with the now-fashionable notion of "active learning." One component here is frequent exercise of skills. For communications ability, identifying opportunities and levels of practice is relatively straightforward, and indeed, numerous associations between the sheer amount of writing and speaking engaged in and growth in these abilities have been documented empirically. Cole (1982) for instance reports many links between frequent in-class writing exercises and speaking in class with growth in these respective abilities--a finding sustained by the second report of the Harvard Assessment Seminar (Light 1992). Using self-reports, moreover, Astin (1993) estimates gains in writing ability and oral communications abilities from freshman to senior year with a number of items noting frequent in-class participation in these activities (e.g., giving presentations in class, taking writing- intensive courses, speaking in class, etc.).

For broader critical thinking and problem-solving abilities, however, the notion of "practice" is more complex. Reviewing the literature on collegiate teaching effectiveness, for instance, McKeachie, Pintrich, Lin and Smith (1986) concluded that three distinct kinds of in-class activities made a difference in promoting thinking skills--student discussion, an explicit emphasis on problem-solving procedures and applications, and stressing the use of "verbalization" and modelling strategies in which students think through a problem. This general finding is sustained particularly by studies of "developmental" courses designed specifically to incorporate such features in the form of journal writing, role-playing, in-class debates, small-group work, or practical problem-solving exercises (Stone 1990; Widick, Knefelkamp, and Parker 1975; Stephenson and Hunt 1977). Additional evidence is provided by correlational studies by Smith (1977, 1981) that among other factors linked student participation in classroom discussions with critical-thinking gain.

A second broadly-sustained component of "active learning" is providing frequent feedback on performance. While this component has been extensively explored in elementary and secondary teaching (e.g., Gagne 1977), most evidence of its effectiveness in collegiate classrooms appears indirect. Direct evidence consists largely of single-course studies with few statistical or physical controls (e.g., Fulkerson and Martin 1981). Across a number of multi-institutional studies, Astin (1985) makes feedback an important principle within his larger concept of "academic involvement" (1977), but most of his evidence rests on self-reports and overall satisfaction with the learning experience. In his later (1993) study of general education outcomes, moreover, "getting frequent feedback" is included in a broader construct called the "humanities orientation" of a given instructional environment, which is also related to self-reported gains. Similarly, in an attempt to relate student perceptions of effective faculty to developmental stage, Baxter Magolda (1987) concluded that faculty efforts to engage students by soliciting comments and reacting to student opinions are effective at all developmental stages. More significantly, she reported that students particularly sought relationships that stressed "partnerships" between faculty and student--a sentiment also reflected in the notions of "coaching" and "cognitive apprenticeship" advanced as effective by Collins, Brown and Newman (1986). These findings are reinforced by the ongoing work of the Harvard Assessment Seminar (Light 1990, 1992), in which students broadly report better learning (but not necessarily higher satisfaction) in courses that provide them with frequent quizzes and other checks on performance.

A third major component of "active learning" is group work and peer interaction. Here there is considerable evidence of positive impact (both direct and indirect) in the development of higher-order thinking skills. Again, some of the most powerful evidence is drawn from non- college sources, but findings appear generalizable to college classrooms (Pintrich 1988; McKeachie, Pintrich, Lin and Smith 1986). Benware and Deci (1984), for instance, found significant differences in higher order thinking skills among high school students who were told that they would be expected to teach others a body of material and a similar group who expected to be tested themselves; interestingly, they found no differences in recall of content between the two groups. Similar findings are reported for college classrooms, but have mostly been confined to gains in content mastery--for example Bargh and Schul (1980) for verbal material or Annis (1983) for historical knowledge. Evidence based on self-reports reinforces these effects. Astin (1993), for instance, reports "tutoring other students" as prominently related to both academic performance based on grades and self-reported gains in communications and higher-order thinking skills; in addition, his summary of results for this longitudinal multi-institutional study places "peer interaction" as one of the three most important factors in explaining growth (together with faculty/student interaction and time on task). Finally, in a series of classroom experiments involving testing and self-reports at Harvard, the effectiveness of small-group work was strongly confirmed (Light 1990, 1992), but particularly in heightening students' enthusiasm and engagement in academic work.

Taken as a whole, this body of evidence for classroom activities related to "active learning" seems both strong and consistent. Because reported effects appear to occur independent of

discipline and context (Pascarella and Terenzini 1991), moreover, developing representative indicators of the presence of these activities appears promising.

3. What Happens in the Wider Environment. Empirical studies also provide substantial evidence that what happens beyond the classroom is important for learning. A considerable tradition of outcomes work, in fact, combines the two in the form of a broad notion of environment or "environmental press" (Feldman and Newcomb 1969, Pace 1979, Chickering 1969). Studies within this tradition rely on both direct cognitive measures and self-reports, and attempt to identify specific characteristics of a collegiate environment that can be linked to cognitive development. In most cases, however, these characteristics appear interdependent and few studies sustain the conclusion that they are individually significant. Consistent across most of them, however, is identification of the selective, liberal arts college as a distinctive exemplar of this environment (e.g., Astin 1977, 1993; Pace 1990, 1984; Winter, McClelland and Stewart 1981; Whitla 1978).

Typical of the specific characteristics noted as part of this environment are those listed by Astin (1993) within the "humanities orientation"--a cluster of factors which he found particularly related to self-reported gains in critical thinking. They include such things as considerable writing, substantial contact with faculty both in and out of class, use of essays in examinations, high levels of participation in class, and an interdisciplinary orientation. Similar factors are identified by Winter, McClelland and Stewart (1981) as part of the "Ivy Experience"--determined by self-report from students in the small private colleges in their sample that showed particular gains in critical thinking skills. They concluded, however, that observed effects were a product of the experience as a whole, rather than being due to any one factor. Using self-reports of both achievement and environment on the College Student Experiences Questionnaire (CSEQ), Pace (1990) identified a distinctive pattern for small, selective liberal arts colleges that emphasized involvement and high participation while at the same time showing unusually high reported gain on such abilities as analysis, synthesis, inquiry, and writing proficiency. Links between the environmental factors reported through the same instrument and learning have also been confirmed using actual achievement data (e.g., Friedlander 1980).

Consistent across most of these studies as well is the finding that faculty-student contact is particularly salient, and is especially prevalent in the liberal arts college environment. Astin's (1993) results, for instance, include a composite factor obtained through faculty surveys that he labels "student orientation", consisting of amounts of faculty time dedicated to teaching and to advising, and amount of reported out-of-class contact with students. Winter, McClelland and Stewart (1981) similarly record "high student-faculty contact" as one of the most prominent features of the "Ivy Experience." As noted earlier, moreover, this environment may contain unusual levels of academic challenge as assessed through the cognitive level of examinations given (Braxton and Nordvall 1985).

Confirmation of the importance of these conditions is also present in other studies not explicitly focused on the benefits of a particular type of college. Pascarella (1989), for

example, identifies out-of- class contact with faculty as one of seven intertwined factors associated with general collegiate benefit on the Watson-Glaser. Similarly, using the institution as the unit of analysis, Centra and Rock (1971) noted high faculty-student interaction as the only environmental feature strongly identified with residual gain on all three components of the GRE general examination after controlling for entering student ability. Student-faculty contact--especially outside the classroom--is the focus of a similar body of work based primarily on student self-report. Associations between reported growth in general cognitive ability have been shown for both the absolute quantity of such contact and with its perceived quality in several longitudinal studies (e.g., Terenzini and Wright 1987; Terenzini, Pascarella and Lorang 1982; Endo and Harpel 1982).

Many of these findings, in turn, can be made part of wider notions of "involvement" in college, reported in the following sections. What seems clear from this pattern of empirical results, however, is that student reports about what happens to them in particular classroom and college environments appear reliably associated with general cognitive gains consistent with the domain of Goal 5.5.

C. Student Behavior

Empirical investigations of the link between what students do and how much they learn in many ways resemble those that examine the effects of classroom activities or the wider environment. Findings on student- faculty contact or on patterns of coursetaking reported previously, for instance, might just as easily be placed under this heading. There is, however, a distinct body of work that centers explicitly on the student's own contribution to the learning process. Findings under this heading can be usefully summarized under two topics--a) time on task, and b) total involvement and student "quality of effort."

1. Time on Task. Not surprisingly, there is a great deal of evidence to support a general association between the amount of time students devote to academic pursuits and the amount of knowledge gained (Pascarella and Terenzini 1991). In general, however, this association is less clear-cut for generic cognitive outcomes than it is for particular subject areas.

Evidence of the impact of time invested is of basically two kinds. The first is indirect, but widely sustained: students gain most in the areas to which they are most exposed and students who enroll for more courses or years of study also gain more. Reviewing the results of multiple "value-added" studies conducted over the past twenty years, Pascarella and Terenzini (1991) confirm Pace's (1979) and Bowen's (1977) conclusions that student performance is enhanced most on those general cognitive dimensions that are closest to their own major fields. DeLisi and Staudt's (1980) results cited earlier are particularly relevant here, as the students they tested on formal reasoning skills did not differ markedly in total score but performed better on problems related to their own field of study. Many studies also demonstrate a relationship between total time invested, as operationalized in terms of years of college study completed, and learning outcomes. Examining NLS data, for instance, Robertshaw and Wolfle (1983) showed differences between two-year and four-year

college verbal and mathematics outcomes, as well as net gains for attending college in general. Finally, there is indirect evidence that students who attend part-time exhibit less gain than those attending full-time. Using self-report data from 59,000 students on a national survey, for instance, Dollar (1992) indicated less reported gain in analytical thinking and communications skills for students attending part-time--though it is important to remember that their perceived starting points may have been different.

A second body of evidence on the impact of varying time investments-- though partially based on self-reports--is provided by studies that attempt to relate time on task with outcomes directly. Astin's (1993) summary of findings on a comprehensive investigation of more than 24,000 students at over 150 four-year institutions concludes after controlling for 170 input variables that student time allocation is the most important single factor associated with college outcomes (closely followed by student-faculty interaction and peer interaction). Examining student responses at multiple institutions, Pace (1990) also reports strong associations between amount of time invested in studying and in academic pursuits, and academic performance. He also reports notable differences between study time at small liberal arts colleges and other types of institutions which are in turn associated with higher self-reported gains in such skills as analysis, synthesis, inquiry, and writing ability. Although far more limited in scope, Johnson and Butts (1983) examined in-class learning outcomes directly, and determined that they were related to the amount of time students spent actively engaged in class activities.

2. Involvement and "Quality of Effort." More significant than simply time invested, moreover, may be the ways time and effort are actually used. Pace (1990, 1984) and his associates (e.g., Friedlander 1980, Porter 1982), for example have conducted numerous analyses associating self- reported skills development in college with a range of "investments" in the learning environment, using the College Student Experiences Questionnaire (CSEQ). In contrast to similar instruments that document student activities while enrolled, this questionnaire is especially designed to solicit items of student experience assumed to be directly associated to learning (Pace 1984). Pace (1990), for instance, constructed a "breadth of experience" index comprising thirteen types of academically-related student activities ranging from course learning and "use of the library," through informal peer topics of conversation, to activities related to student clubs or other organizations. Reporting on the responses of over 10,000 students at 33 institutions, he reports year-by-year gains in such skills as analysis, synthesis, inquiry, and writing, and associates reported gains strongly with scores on the breadth index; in each skill area, differences of more than fifty percent were obtained between low- breadth and high-breadth respondents. Friedlander (1980) associated actual college performance with many of these factors, although the breadth index itself was not used. Also using self-reported levels of involvement, Terenzini and Wright (1987) found significant associations between reported development and amount of involvement in a variety of learning activities.

At the classroom unit of analysis, similar findings have been reported (McKeachie, Pintrich, Lin and Smith 1986). Pintrich (1988), for example, has used student self-reports in such

areas as motivation to learn and studying behavior to investigate learning gains in particular classes. Using the Motivated Strategies for Learning Questionnaire (MSLQ), Pintrich and Johnson (1990) report significant associations between high scores and student outcomes. As noted earlier, these findings are consistent with those of Johnson and Butts (1983) on the actual utilization of student time in the classroom.

Astin (1977, 1985, 1993) and others have carried this notion farther by positing a more general theory of "student involvement" as an explanation of collegiate learning and development. According to Astin (1985) particular areas of engagement or activity elements matter far less than the aggregate amount of "psychological energy" that a student invests in the experience. A parallel implication is that individual factors may be impossible to differentiate empirically; what matters is the "total environment created by students and faculty" (Astin 1993). Astin's (1993) multi- institutional study of general education provides considerable evidence for this contention--although like most prior work, stronger associations are shown between involvement measures and self-reported gains than when using actual cognitive outcomes measures. Dimensions of involvement reported tend to confirm Pace's (1984) position that academically-relevant (but not necessarily in-class) activities are important. For instance, critical thinking gains are seen as strongly associated with the "humanities orientation"--a construct that includes such factors as considerable writing activity, faculty involvement in teaching general education courses, essay assignments, interdisciplinary classwork, participation in class, and out-of-class student-faculty contact.

Other investigations, as noted, support the notion that many elements of student activity interact to create a combined effect. These include Winter, McClelland and Stewart's (1981) "Ivy Experience," studies that employ Tinto's (1975) notions of "academic and social integration" to examine self-reported learning gains (e.g., Terenzini, Pascarella and Lorang 1982; Terenzini and Wright 1987), and direct examinations of gains in critical thinking. In one of the few true causal studies of this nature, for instance, Pascarella (1989) found seven elements of the collegiate experience related to critical thinking gains--including living on campus, time on task (hours spent studying), out-of-class contact with faculty, peer contact, attendance at lectures and debates, amount of unassigned reading, and extra- curricular involvement--but could not dissociate the factors statistically.

Taken together, this body of evidence supports the contention that student activities and levels of involvement are strongly related to the development of general collegiate abilities. At least as importantly, this literature demonstrates that reports about such activities obtained directly from students appear consistent, and can be obtained relatively straightforwardly. Although few empirical studies can demonstrate causal links between "involvement" and outcomes (and indeed, arguments can easily be sustained that the two are mutually reinforcing [e.g., Pascarella and Terenzini 1991]), associations appear sufficiently robust in this arena to yield considerable confidence in a potential indicator.

D. The Special Case of Student Self-Reports

A large proportion of the empirical literature that attempts to link learning gains with other characteristics of students and the learning environment is founded principally upon the use of self-reported data (e.g., Astin 1977, 1993; Pace 1984, 1990; Terenzini and Wright 1987; Endo and Harpel 1982). Student responses to questionnaire items have in each of these cases been used to operationalize both cognitive and environmental factors, which are then associated with one another. From the perspective of developing useful indicators consistent with Goal 5.5, this experience with self-reports raises two distinct special issues--both related to the ultimate validity of self-reports. First, if reasonable validity is established, findings that link environmental characteristics and student experiences with outcomes can be used to support the case for constructing indicators based on these characteristics and experiences--however these are assessed. It is precisely because such a case for self- report validity can be made that findings of this kind have been included in this review as, indeed, they generally are in the scholarly literature (Pascarella and Terenzini 1991). Second, however, if the validity of student self-reports about cognitive gains can be established, they can themselves serve as a "proxy" measure of actual knowledge gain--and be usefully included in a national indicators system.

There is a considerable literature concerned with establishing the validity of student self-reports about cognitive outcomes. Arguments here are generally of two types. One line of work attempts to link self- reports directly to examination-based results through cross-testing the same population. Baird (1976), for instance, conducted an extensive review of primarily course-based studies that compared self-reported knowledge gains with actual outcomes. His results suggest that the two do indeed vary together dependably, though they are far from coincident. Dumont and Troelstrup (1980), using the ACT-COMP examination, administered parallel self-report items and also found substantial agreement between the two on such general cognitive outcomes as "critical thinking ability" and "knowledge of different methods of inquiry;" but on the basis of their results, they cautioned against using self-reports alone as measures of these attributes and advocated a multiple indicators approach. Using national survey data from the CIRP and matched examination results on the GRE and LSAT examinations Anaya (1992) also found positive correlations for general abilities. Such results suggest the conclusion that self- reports of cognitive gain are indicative of, but not completely coincident with, results obtained through more direct forms of assessment.

The second line of argument for the validity of self-reports concerns the degree to which patterns of results obtained in this manner parallel those produced by more direct cognitive measures. Pace (1990), for instance, provides five reasons to support the validity of such measures including, a) the consistency of results over time and across different populations, b) the fact that patterns of outcomes vary for self-reports across majors and length of study in the same manner as has been established through direct achievement testing, c) the internal consistency of questionnaire responses across different items on the same dimension, d) the fact that reported growth follows patterns of experience that should be expected, and e)

apparent student seriousness and engagement with the instrument itself. Of these arguments, the second is probably the most convincing, and is widely supported in other studies. Astin's (1977, 1985, 1993) results on over twenty years of experience with the CIRP Freshman and Follow-Up surveys show patterns of self-reported outcomes that vary consistently by major field and other measures of levels of exposure, just as directly-assessed cognitive outcomes do. Similar patterns, moreover, have been reported by alumni on national samples (Dollar 1992). Cohen's (1981) meta-analysis of student course evaluations also obtained considerable consistency of results in student ratings of course characteristics expected to contribute to learning gain and actual levels of class achievement. An important caution in many of these cases, however, is that internal relationships among questionnaire items addressing cognitive and environmental factors that are administered together virtually always show a stronger relationship in such studies than when cognitive instruments are used. The bulk of Astin's (1993) findings regarding associations between involvement factors and learning, for instance, are based on relationships among self-reported items, not results on the GRE and LSAT which he also examined.

Overall, the bulk of evidence available on this matter appears to support Dumont and Troelstrup's (1980) confirmation of the utility of self-reports and their call for multiple measures. Self-reported data on both cognitive attainment and on instructional "good practices" or other environmental factors can easily be collected on large samples; and based on past research experience, results obtained by this method will be consistent with more direct measures. But carefully-controlled research studies based on more direct forms of assessment will also be needed to empirically anchor this approach.

IV. Review of Available National Data-Gathering Instruments and Approaches

Data sources potentially useful for generating indicators of good practice are of many kinds. Appropriate instruments and methods have been extensively used at the institutional level and occasionally at the state level, but few are currently in place at the national level. Modifying many of these tools for use with national samples, however, would not be technically difficult, and could be done reasonably efficiently and quickly.

Results of the literature review that identified potentially-promising instruments, methods, or databases for assembling "good practice" indicators are summarized in this section. Like the identification of principal empirical findings, those included were subject to a number of caveats. First, of course, those mechanisms identified had to address one or more of the dimensions of "good practice" discussed in the previous section. Second, those mechanisms identified had to be sufficiently proven on technical grounds that assurance could be placed in their ability to generate information reliably. Third, those mechanisms identified had to be national or potentially national in scope; that is, their features and design should be sufficiently general that these instruments could be suitably deployed across a wide range of postsecondary settings and types of students. Application of these three criteria resulted in a relatively short list of potential data- gathering tools, grouped under a number of categories. These included, a) aggregate statistical reporting systems, b) surveys or inventories of

institutional practices, c) methodologies for conducting student transcript studies or coursetaking analyses, d) surveys of faculty teaching practice, and e) surveys of current or graduating students. Instruments or approaches of particular promise for further exploration are noted under each of these headings, and their strengths and weaknesses briefly assessed.

A. Institutional Administrative Records

Colleges and universities currently keep extensive records on instructional activity and resource utilization. If standard indicator definitions could be developed, these data might potentially be tapped to help determine how institutions are deploying available resources in support of effective undergraduate education. Examples of the kinds of indicators that might be produced from statistics of this kind include, average class sizes in key classes, proportion of lower-division classes taught by graduate students and full-time faculty, and the proportion of small classes typically taken by first-year students. Such data might be collected on the basis of a national cross-sectional sample of institutions. Alternatively, key statistics might eventually be made a part of institutional reporting to the National Center for Education Statistics (NCES) under the Integrated Postsecondary Education Data System (IPEDS).

At present, however, no established national methodologies for compiling such statistics exist. Their closest approximations at the national level are noted and assessed below.

1. The Integrated Postsecondary Education Data System (IPEDS). NCES currently collects a range of descriptive information from all postsecondary institutions on a regular basis. These might, in turn, be used to construct indirect measures of institutional practices through the calculation of appropriate ratios among base measures. Currently, ratios or performance measures based on available data in IPEDS that might be relevant include:

- o headcount and FTE enrollments (including first-time enrollments) by number of faculty by rank, and tenure status.

- o instructional expenditures per FTE student.

Neither of these alternatives will likely yield data of sufficient validity to serve as an appropriate indicator of good practice.

2. Existing State Indicators Systems. In response to growing demands for accountability for public institutions, some fifteen states have developed performance indicators for public reporting. The majority of these have been implemented only recently, and the quality of the data being compiled is as yet unknown. Among the most relevant indicators contained in such systems are the following:

- o proportion of resources devoted to undergraduate instruction (e.g., New York-SUNY)

- o numbers and proportion of full-time faculty engaged in teaching freshmen (e.g., Kentucky).

- o numbers and proportions of undergraduate students participating in faculty research activities (e.g., South Carolina).

- o numbers of students receiving a "capstone" or comparable integrative experience at the end of their period of study (Minnesota State University System).

- o student satisfaction as assessed by survey (e.g., Tennessee, New York-SUNY).

Although marginally more relevant than the overall ratios derivable from IPEDS, such macro-level indicators are similarly limited-- especially in the light of findings of the research literature that experiences in particular "micro-settings" (such as the classroom) are principally associated with cognitive attainment. NCES should examine the experiences of these states closely, however, to determine the potential of such indicators; in particular, any state-level attempts to relate progress on one or more of these indicators to a state-level assessment of cognitive skills would be worth tracking and reporting.

B. Surveys of Institutional Practice

Institutional surveys have in the past been used by higher education scholars to determine the degree to which colleges and universities are engaged in innovative practices in undergraduate instruction. Prominent examples here include a study on undergraduate reform recently undertaken by the National Center for Research on Postsecondary Teaching and Learning (NCRPTAL) at the University of Michigan (Peterson 1987) and an ongoing "Registry of Undergraduate Reform" maintained by the California State University System (Vandament 1991). An available mechanism for conducting such surveys, moreover, is provided by the "quick-response survey" capability maintained by NCES. Such instruments typically rely on "expert" respondents at each institution, generally selected by position, to report on institutional activities, and generally a sample of institutions is surveyed. Examples of the kinds of indicators that might be obtained by means of this method include minimal skills and curriculum- coverage requirements for receipt of the baccalaureate, particular curricular emphases on critical thinking, communication and problem- solving, and institutional activities in the assessment of student learning.

No currently-established survey contains all items of relevance for the development of a proposed indicator. Several, however, contain some such items and have proven the feasibility and utility of this approach.

1. Academic Management Practices Survey. This survey was developed by researchers at the National Center for Research to Improve Postsecondary Teaching and Learning (NCRPTAL) at the University of Michigan (Peterson 1987), and was designed to assess

institutional "academic management practices" thought to be consistent with effectiveness in undergraduate teaching and learning. Items were developed through a literature review of this topic, and emphasize practices suggested by the literature to be effective. The survey concentrates on two main issues, a) the identification of problems that inhibit the improvement of undergraduate teaching and learning and, b) specific academic management practices. For the latter, "expert" respondents are requested to indicate whether these practices exist, whether they are newly developed, and the degree to which they have proven effective. Items are predominantly concerned with academic policies and the existence of particular kinds of resources (e.g., instructional development, technology, etc.). Few items address actual classroom delivery--but a number examine the incentive system in place to support instructional innovation and faculty investment in undergraduate education. The survey has been administered successfully to representatives of over 400 four-year colleges and universities.

2. Seven Principles for Good Practice--Institutional Inventory. This instrument was developed to assist institutional leaders in assessing the degree to which current academic practices are consistent with the Seven Principles (Chickering and Gamson 1987, Gamson and Poulsen 1989). Although intended principally as a "reflective device" to provoke thinking at the policy level, the inventory contains a number of items directly consistent with areas that have been documented empirically as effective in promoting cognitive gain. Items on learning communities, field experiences and applied learning, mastery learning, and interdisciplinary opportunities in the curricular area; or on advisement and faculty incentives in the faculty area, are particular appropriate for further development in the light of evidence from the research literature.

Overall, this line of development appears promising, but with several important caveats. First, the use of self-reports from institutional leaders raises significant issues of credibility--particularly in a high- visibility environment. Second, overall institutional surveys--especially for large institutions--may miss the most important practices that are present or absent at the unit or department level. For these reasons, it appears unwise to invest too heavily in developing institutional inventories of this kind for use in a national indicators system unless these are anchored by other measures. On the other hand, such inventories might provide excellent examples of the kinds of items to include in such occasional or supplementary data-gathering efforts as NCES "Quick- Response" surveys.

C. National Transcript Studies

Transcript records contain information on typical college coursetaking patterns, and can be used to determine the degree to which students are generally exposed to particular bodies of material. Several national studies of this kind have been undertaken for different purposes, and a standard coding scheme has been developed to categorize course data (Adelman 1990). Such studies usually examine coursetaking patterns on a discipline basis--addressing questions such as the number and proportion of courses in a total baccalaureate career taken in key identified areas (for instance, science, math, or history), or the overall "concentration" and

"cohesiveness" of the curriculum. Though challenging, similar methodologies based on a national sample might be developed that, together with supplied course descriptions and required assignments, could suggest the frequency with which students are typically graduating after having completed certain key assignments or experiences--for example an independent research project or a senior "capstone" experience. Finally, consistent with a transcript analysis, actual examples of student assignments or examinations might be analyzed for cognitive content using methodologies such as those suggested by Pintrich (1988) and Braxton and Nordvall (1985).

Two primary methodologies for examining transcripts have been used on multiple institutions that provide useful models of this approach.

1. Structure and Coherence in the Undergraduate Curriculum. This curriculum/coursetaking methodology was developed by Zemsky (1989) and his colleagues at the University of Pennsylvania in association with the Association of American Colleges (AAC). Its principal purpose is to empirically examine patterns of coursetaking as noted in graduating senior transcripts by discipline and type of institution. An additional data source is catalogue material that indicates the courses available in each discipline, and their prerequisite sequences. Originally piloted on 30 four-year institutions, the national database has been gradually expanded to include over 75 institutions. Using standard course classifications, the method constructs a "breadth" index that indicates overall levels of student exposure to particular disciplinary groupings, and a "depth" index that indicates the proportion of courses available in a given domain that are part of a structured sequence based on prerequisites. Results have not been linked in any way with cognitive attainment or other outcomes measures, but have produced striking descriptive patterns of different kinds of curricular structures.

2. The Coursework Cluster Analysis Model (CCAM). This method was developed by Ratcliff and Associates (1988; Ratcliff 1993) to identify combinations and sequences of courses, derived from transcript data, that are associated with gains in general education outcomes. The method is intended to be used directly in combination with measures of learning outcomes, and was piloted on GRE general examination results residualized by entering SAT score. In essence, it uses statistical clustering algorithms to identify particular types and sequences of courses associated with particular outcomes. The method has been used extensively in analyses of six institutions to determine the effectiveness of general education, and in research on curricular structure (e.g., Jones and Ratcliff 1991). Results have shown that the method is effective in differentiating identifiable patterns of coursework that are empirically related to cognitive gain.

Transcript-based approaches such as these appear particularly promising for generating useful indicators of institutional practice. It must be remembered, however, that most empirical studies suggest only relatively modest associations between structural aspects of the curriculum and general cognitive outcomes. Combined with available outcomes data, such

methods do appear particularly useful in examining trends in an arena where policy or institutional action is unusually visible and direct.

D. Surveys of Faculty Teaching Practice

Higher education researchers have also developed many surveys of college and university faculty. While most are directed at issues of career, compensation, and scholarly research, several also contain items on teaching practice. Examples of the kinds of indicators that might be developed using this method are the proportion of faculty time spent working with students, the reported incidence of "active learning" activities conducted by faculty members, levels of student participation in independent or faculty research work reported by faculty, and levels of faculty participation in professional development activities directed toward the improvement of teaching. Surveys of this kind can also use faculty as "expert witnesses" regarding typical institutional practices, outcomes, and reward structures. Faculty surveys, for instance, often ask respondents to rate such items as the frequency of student-faculty contact, the importance in instruction of various intended outcomes of college, and the typical types of evaluation methods used to assess student classroom performance.

Two faculty instruments or inventories have been developed that are especially relevant to the identification of good practice in teaching and in the supportiveness of the wider institutional environment.

1. UCLA Faculty Survey. This instrument was developed by Astin and Associates (Astin, Dey, and Korn 1991) for use in conjunction with national studies of college student development conducted through the Cooperative Institutional Research Program at University of California, Los Angeles. Results have been used in conjunction with other data to investigate factors associated with student cognitive development in college (e.g., Astin 1993). Items on the questionnaire directly related to good practice include, a) information about faculty behavior such as teaching techniques and types of examinations administered, and b) information about faculty perceptions of the institutional environment such as support for teaching or student support. Particularly appropriate for identifying good practice--and empirically related to a number of measures of student outcomes--are items on the use of such instructional techniques as cooperative learning, student presentations in class, experiential learning, peer evaluation, independent projects, class discussions, or student-developed assignments. Environmental and values factors especially relevant to this topic include ratings of the importance of enhancing students' self-understanding, enhancing out-of-class contacts and experiences, and the degree of student peer interaction. Items such as these are combined to yield a number of scales intended to characterize particular institutional environments (for example, the "humanities orientation" or "student orientation"). These can then be related to specific institution-level measures of student outcomes.

2. Seven Principles for Good Practice--Faculty Inventory. This instrument was developed in parallel with the Institutional Inventory described previously to help guide faculty in

conducting self- assessments of their own practices consistent with the Seven Principles for Good Practice (Chickering and Gamson 1987, Gamson and Poulsen 1989). Items in the inventory ask faculty members to rate the frequency with which they engage in specific instructional activities consistent with each of the seven principles, and include items on both classroom techniques and behavior, and more general contact with students. Like the parallel Institutional Inventory, specific items included on the Faculty Inventory were selected on the basis of empirical research evidence of their association with student learning and development. While systematic research linking responses on the instrument to specific institutional characteristics or educational results is lacking (and indeed, it was never the intent of the instrument to be used as a research tool per se), the activities noted constitute perhaps the best single list of such items available for administration to faculty. Many could also be easily adapted for inclusion in a student questionnaire.

The potential of such instruments to add to an understanding of student learning and development has been directly demonstrated (e.g., Astin 1993, Angelo and Cross 1993). While questions can certainly be raised about the validity of faculty self-reports in these areas, most results reported are consistent with what is known about faculty from other sources (Astin, Dey and Korn 1991). Similarly, unlike the administrative leaders who typically complete the institutional surveys described earlier, line faculty have little at stake in such surveys, and are thus more likely to provide an objective response. Available vehicles for administering the required questions to valid national samples of faculty already exist in the form of NCES' National Survey of Postsecondary Faculty (NSOPF), that already contains items on workloads, courses taught, and attitudes.

E. Surveys of Current and Graduating Students

Questionnaires administered to current and former college students are typically used by individual colleges and universities to determine levels of satisfaction with instruction and other services, patterns of typical student activities, and self-reported outcomes of instruction. They also constitute a significant data source for empirical research on college impact. Several of these questionnaires have been administered on a national basis. Items contained in these surveys include reported levels of participation in activities such as internships or faculty research projects, time spent in various activities (for example, group study or tutoring another student), frequency of student-faculty and group interaction, self-reported class content. Most also contain items on self-reported gains on a range of outcomes. Many of these items are potentially usable as indicators without change; for the future, enhancements might be made in the item content and in the sampling base used for such instruments to render them suitable for inclusion in a national indicators system.

A number of current student survey instruments have been used widely to investigate college student experiences and behavior, and contain items of interest.

1. Learning and Study Strategies Inventory (LASSI). This instrument was developed by researchers at the University of Texas, Austin (Weinstein, Schulte and Palmer 1987) to

assess individual student approaches to particular kinds of learning tasks. Items request students to rate the degree to which particular activities are "typical" of the way they work, and are combined to yield a number of scales--including motivation, anxiety, time-management, or concentration. Scales are also produced on particular strategies of learning employed that are consistent with the tenets of active learning--for example, self-testing, selecting main ideas, constructing and using study aids, or test strategies.

2. Motivated Strategies for Learning Questionnaire (MSLQ). This instrument is similar to the LASSI, and was developed by Pintrich and Associates at the University of Michigan (Pintrich 1988; Pintrich, McKeachie and Smith 1989). Scales are also based on self-reported items that address student beliefs about themselves and how they work; these scales are centered on three main topics--general motivation, need for extrinsic reward (such as grades, or approval), and the intrinsic rewards associated with learning. Research findings suggest that effective students are high on all three dimensions.

3. Cooperative Institutional Research Program (CIRP) Surveys. These are comprehensive surveys that have been in use for over twenty-five years, developed by Astin and Associates at the University of California, Los Angeles (e.g., Astin and Associates 1992). The Freshman Survey is administered widely by individual colleges and universities, and data on a representative national sample is also collected. The Follow-Up Survey is administered to a systematic sample of students previously completing the Freshman Survey and is used for longitudinal research purposes. Both surveys contain a range of self-reported items on typical activities, goals, and self-ratings. The Follow-Up Survey also contains additional items on collegiate experience, and items on self-reported development. Items on experience are broad, but among them are questions that address contact with faculty, peer tutoring, class attendance, and outside academically-related pursuits. Items on self-reported knowledge include several related to Goal 5.5 outcomes including critical thinking and communications ability.

These surveys are currently administered on a regular basis to students at over 400 colleges and universities and have been used in a number of prominent, multi-institutional investigations of collegiate outcomes (Astin 1977, 1993). Experience and environment items have been linked to GRE and LSAT residual score results after controlling for SAT (Astin 1993), and self-report items have been validated by empirical linkages to the results of direct cognitive assessment (Anaya 1992).

4. College Student Experience Questionnaire (CSEQ). The CSEQ was developed by Pace at the University of California, Los Angeles (1984, 1987), specifically for use in researching academically-related activities and experiences expected to have an impact on learning. The instrument is based on the notions of "investment" and "quality of effort", and items center on specific levels of involvement associated with particular academic resources or features of the environment. Items are combined into a number of Activity and Quality of Effort Scales that exhibit excellent psychometric properties (Pace 1987). The CSEQ also contains a number of items on self-reported growth-- including questions on "inquiry, synthesis, analysis, and

communication" skills consistent with the domain of Goal 5.5. The CSEQ has been used at over 150 colleges and universities, and national data have been compiled for comparison studies. Activity and Quality of Effort Scales have been used in a number of studies of collegiate impact (e.g., Pace 1990), and have been linked to academic performance (Friedlander 1980).

5. ACT-ESS Alumni Questionnaire. Available from the American College Testing Program (ACT 1982) as part of the Evaluation Survey Service (ESS) of twelve questionnaires, the Alumni Survey has been administered since 1979 to graduates of a variety of colleges and universities. The instrument contains a number of self-report items on cognitive skills dimensions consistent with the domain of Goal 5.5, and has been used in empirical studies of college impact (Dollar 1992) and in several state and system-wide reporting systems.

6. NCHEMS SOIS Questionnaires. Developed by the National Center for Higher Education Management Systems (NCHEMS) in conjunction with the College Board (NCHEMS 1983), the Student Outcomes Information Service (SOIS) questionnaires are intended to provided college and university administrators with evaluative information on the effectiveness of programs and services. Six instruments are available--ranging from an entering student survey, through a continuing and former student survey, through completer and alumni surveys. All six surveys contain a common core of self-reported goal items, and ask students to address the degree to which each goal is important to them and the degree to which the institution has enhanced their attainment of the goal. Several of these goal items are related to the domain of Goal 5.5.

Questionnaires of this type have already been widely used to collect information relevant to a national good practices indicator system. As noted in Section III, items on student experiences and levels of involvement have been linked extensively to cognitive outcomes. Instruments such as the CSEQ and CIRP surveys have been proven on national samples, and provide an efficient means to collect data from a wide range of students and institutional settings. For a number of these instruments, moreover, historical norms are available that can be broken down by types of institutions and student clienteles to track progress. As in the case of faculty surveys, moreover, available national vehicles for administering the types of items required may already exist in the form of NCES' Recent College Graduate Study, soon to be established longitudinally as "Baccalaureate and Beyond."

As in Section III, it is not the intent of this section to examine every existing survey or data-gathering methodology used in prior investigations of the factors related to collegiate learning. Rather, the objective is to identify particularly promising types of evidence-gathering, and to describe relevant existing instruments or methods that are readily available and that have been administered successfully across many different kinds of institutions and settings. As a final note, it should be recognized that a diversity of data sources is itself important in developing a reliable set of indicators. Indeed, the nature of indirect indicators is such that confidence in the message conveyed is as much a product of the number of very distinct data sources tapped as it results from the technical precision or

validity of any one method. Many of the same types of indicators can and should be collected by means of several different sources.

V. Summary and Recommended Actions

It is important to re-emphasize that this review was undertaken for a limited set of purposes. First, effective policy demands obtaining information about all parts of the higher education enterprise on a consistent basis--including inputs, resources, processes, and outcomes. Good practice indicators were thus always intended to supplement data gathered by other means. To be effective in guiding policy, however, the extent of their validation as good practices needed to be clearly established. One purpose of the review, therefore, was to summarize what is known about the empirical connections between such practices and the collegiate outcomes identified in Goal 5.5. Second, the policy context demands the development of relevant, useful data that can be quickly and efficiently generated on a national basis. Indicators of good practice were initially proposed in this environment partly because a number of methods for collecting them were already developed. Another purpose of the review, therefore, was to surface the leading candidates currently available as models or as vehicles for assembling the needed data. Third and finally, the policy context favors the development of indicators that are in some way linked to policy action. Although both outcomes and "proxy" indicators may be useful in documenting the existence of a condition, they provide little information about what ought to be done. A final purpose of the review, therefore, was to examine the most promising candidates for "good practice" indicators in the light of their ability to inform appropriate policy action.

Findings of the review on these dimensions are summarized in the accompanying chart. Entries under the domain dimensions column of this chart are consistent with subsection headings of Section III, and describe particular types of potential indicators. Additional columns of the summary chart contain the following:

- "Relative Strength" provides an overall rating of the association between each domain or dimension and general collegiate attainment based on the empirical literature, as reported in Section III. This rating is judgmental, and is based on the overall weight of evidence reviewed. As a result, it is far from definitive. But the results of this exercise appear sufficiently strongly-patterned that they suggest some conclusions. Available evidence, for instance, suggests far stronger linkages between desired outcomes and what happens in class and what students do, than it suggests the importance of particular institutional investments or curricular structures. Where the latter are important, moreover, it appears to be largely because of the opportunities for actions and behaviors that they enable, and not so much because these features are important in themselves.

- "Available Methods" summarizes the primary methods in place for collecting data on each domain or dimension, as described in Section IV. Again, many methods might

potentially be included here that were not explicitly reviewed; those included represent approaches, a) that are currently available, b) that contain items that provide a reasonable match with the domain in question, c) that have been used across multiple types of institutions and settings, and d) for which there is a reasonable body of experience in implementation and reasonable evidence of construct validity. It is reassuring to note that at least some measures or approaches that meet these tests are available for each potential indicator domain or dimension.

- "Relative Ease of Data-Gathering" provides an overall rating of the level of investment that would be needed to collect information on each domain or dimension, using one or more of the methods noted in the previous column. This is also a judgment-based rating, but it is derived from a range of past experience with the use of such methods. Accordingly, national questionnaire studies based on existing instruments and approaches are rated as far less difficult to implement than such methods as catalogue reviews or ratings of the difficulty of student assignments that require extensive assembly of materials and the application of sophisticated methodologies.

- "Policy Relevance" provides a judgment-based rating of the degree to which information on each domain or dimension can be directly related to policy decisions. High "relevance" in this case means that the domain is clearly related to decisions that can be made or conditions that can be immediately affected at the policy level, while a low rating indicates that the domain in question is less likely to be quickly addressed. Ratings of high relevance, for instance, are assigned to such areas as curriculum structure or class size because such factors can be directly affected by institutional policy or resource allocation. Only moderate ratings are assigned to more complex factors that will require changes in teaching practice at the classroom level, in faculty and student behavior, or in the wider climate for teaching provided by the institution--even though these factors have been shown to be far more strongly associated with student learning.

The "Overall Potential" column of the chart provides a summary rating of the assessed potential of each domain or dimension for development as part of a national indicators system, based on the review as a whole. The review's primary conclusion is that the greatest potential for the development of indicators of good practice consistent with Goal 5.5 appears to lie with questionnaire-based data collected around specific instructional good practices, the degree of academic involvement and student-faculty contact present in the wider institutional environment, and specific student behaviors related to time on task and "quality of effort." This approach would rely on both student and faculty surveys, administered to systematic national samples. Also recommended for development, but more difficult to accomplish, are indicators related to the "behavioral curriculum" based on available national transcript methodologies, and parallel studies of the levels of difficulty of representative cross-sectional samples of collegiate assignments and examinations. A final

recommended line of development is the inclusion of appropriate items on self-reported gain on any student questionnaires to be administered.

Given these recommendations, an appropriate next step would be to conduct a feasibility study to determine the costs and logistics associated with collecting information of this kind from a typical sample of institutions and potential respondents. While in some cases, instruments and methodologies for collecting items of interest have been developed and used on national samples, in most cases they have not. Accordingly, the major goals of a feasibility study should be to develop an appropriate set of draft data collection instruments consistent with the above recommendations, including faculty surveys, student surveys, and transcript/student assignment-coding methodologies, and then pilot these approaches on a reasonable sample of institutions or respondents.

Alternatives for such a study might involve, a) proceeding in conjunction with an ongoing national study (for instance, the longitudinal study currently being conducted by the National Center on Postsecondary Teaching, Learning and Assessment at the Pennsylvania State University), b) selecting all institutions in a single state that already possesses a statewide reporting system for higher education outcomes (for example, Florida or Tennessee), or c) selecting a representative panel of institutions that have already conducted appropriate cognitive assessment activities and/or have had extensive experience in administering longitudinal studies. Objectives of the pilot would be to both assess the feasibility of collecting a range of good practice measures themselves, and to further establish empirical links between obtained indicators data and available data on postsecondary outcomes at each institution. Such an approach would likely require a year-long development effort before a final set of appropriate national indicators of good practice could be fielded. But this timeline is far shorter than that required to develop a direct assessment of collegiate attainment.

It is important to recall that the case for developing indicators of good practice consistent with Goal 5.5 rested originally on two grounds. First, direct assessments of the ability of college graduates to "think critically, communicate effectively and solve problems" are technically daunting and will be a long time in coming. Less direct approaches raised the promise of generating useful data at an earlier point, allowing postsecondary education to remain at the center of national interest. Second, the information provided by such indicators would be of utility in itself in guiding the development of national policy. If such indicators could be reliably linked to desired outcomes, they might be of far greater value in inducing institutional change than would outcomes indicators used alone.

Results of this analysis suggest that this original case is justified. Appropriate indicators of good practice that can be empirically related to desired collegiate outcomes are feasible, and the technology needed to implement them is available. Pursuing this path in conjunction with the development of national assessments of these abilities appears an efficient and effective direction for policy.

Summary Chart of Potential "Good Practice" Indicators

Indicator Domain/Dimension	Relative Strength of Association with Goal 5.5 Outcomes	Available Methods for Collecting Indicators Data	Relative Ease of Data-Gathering	Policy Relevance	Overall Potential
A. Institutional Requirements:					
Broad General Education Curriculum/Requirements	Weak/Moderate	- Catalogue Review - Institutional Questionnaires/ Inventories (e.g., Peterson 1987)	Difficult	High	Low
Breadth of Coursetaking/ Types of Coursetaking	Moderate	- "Breadth" and "Depth" of Courses Taken (Zemsky 1989) - "Differential Coursework Methodology" (Ratcliff & Associates 1988)	Moderately Difficult	Moderate	Moderate
Special Courses/ Instructional Designs	Moderate (but likely derivative)	- Catalogue Review - Institutional Questionnaires/ Inventories (e.g., Gamson & Poulsen 1989)	Difficult	Low	Low
Levels of Expectation	Moderate/Strong (via association with "Small Liberal Arts College" effect)	- rating examinations and course materials by level of difficulty (e.g., Braxton & Nordvall 1985)	Difficult	Moderate	Moderate

Indicator Domain/Dimension	Relative Strength of Association with Goal 5.5 Outcomes	Available Methods for Collecting Indicators Data	Relative Ease of Data-Gathering	Policy Relevance	Overall Potential
B. Instructional "Good Practice":					
Class Size and Structure	Low/Moderate (but likely derivative)	- Institutional Surveys/Statistics	Moderately difficult	High	Moderate
"Active Learning" practices in class		- Faculty Surveys e.g., 7 Principles Surveys (Gamson & Poulsen 1989),	Not difficult	Moderate	High
- Practice of Skills	Strong	UCLA Faculty Survey			
- Frequent Feedback	Strong				
- Peer Interaction	Strong	- Student Surveys e.g., CBEQ (Pace 1987) CIRP (Astin & Associates 1992)			
Wider Institutional Environment					
- Involvement	Strong/moderate (complex interaction of factors)	- CSEQ (Pace 1987) - CIRP (Astin & Associates 1992)	Not difficult	Moderate	High
- Student/Faculty Contact	Strong/Moderate (but also may require data on nature of interaction)		Not difficult	Moderate	High

Indicator Domain/Dimension	Relative Strength of Association with Goal 5.5 Outcomes	Available Methods for Collecting Indicators Data	Relative Ease of Data-Gathering	Policy Relevance	Overall Potential
C. Student Behavior:					
Time on Task	Strong/Moderate (but also requires quality of investment measures)	- CSEQ (Pace 1987) - CIRP (Astin & Associates 1992)	Not difficult	Moderate	High
"Quality of Effort"/ Involvement and Investment	Strong	- CSEQ (Pace 1987)	Not difficult	Moderate	High
D. Self-Reported Cognitive Development:					
	Moderate/Strong	- CSEQ (Pace 1987) - CIRP (Astin & Associates 1992) - ACT-ESS (ACT 1982) - NCHEMS SOIS (NCHEMS 1983)	Not difficult	N/A	Moderate

REFERENCES CITED IN TEXT

Adelman, C. (1990). *A college course map: Taxonomy and transcript data*. Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education.

Allred, G.H. and Graff, T.T. (1980). Improving students' interpersonal communication. *Journal of College Student Personnel* (March), 155-162.

American College Testing Program (1982). *The ACT evaluation/survey service: User's guide*. Iowa City: The American College Testing Program.

Anaya, G. (1992). Cognitive development among college undergraduates. Doctoral Dissertation. Los Angeles: University of California, Los Angeles.

Angelo, T.A. and Cross, K.P (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco: Jossey-Bass.

Annis, L.F. (1983). The processes and effects of peer tutoring. *Human Learning*, 2, 39-47.

Annis, L.F. and Annis, D.B. (1979). The impact of philosophy on students' critical thinking ability. *Contemporary Educational Psychology*, 4, 219-226.

Association of American Colleges (1985). *Integrity in the college curriculum: A report to the academic community*. Washington, DC: Association of American Colleges.

Astin, A.W. (1968). Undergraduate achievement and institutional "excellence." *Science*, 161, 661-668.

Astin, A.W. (1977). *Four critical years*. San Francisco: Jossey-Bass.

Astin, A.W. (1985). *Achieving educational excellence*. San Francisco: Jossey- Bass.

Astin, A.W. (1993). *What matters in college?* San Francisco: Jossey-Bass.

Astin, A.W. and Associates (1992). *Cooperative Institutional Research Program (CIRP)*. Los Angeles: Higher Education Research Institute, The University of California, Los Angeles.

Astin, A.W., Dey, E.L., and Korn, W.S. (1991). *The American college teacher*. Los Angeles: Higher Education Research Institute, the University of California, Los Angeles.

Baird, L.A. (1976). Structuring the environment to improve outcomes. In O.T. Lenning (Ed.), *Improving educational outcomes*. (New Directions for Higher Education, No. 16, pp. 1-24). San Francisco: Jossey-Bass.

Banta, T.W. (1986). *Performance funding in higher education: A critical analysis of Tennessee's experience*. Boulder, CO: National Center for Higher Education Management Systems.

Banta, T.W., Lambert, E.W., Pike, G.R., Schmidhammer, J.L., and Schneider, J.A. (1987). Estimated score gain on the ACT COMP exam: Valid tool for institutional assessment? *Research in Higher Education*, 27, 195-217.

Banta, T.W., and Pike, G.R. (1989). Methods for comparing outcomes assessment instruments. *Research in Higher Education*, 30, 455-470.

Bargh, J., and Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, 72, 593-604.

Barnes, C.A., Ed. (1992). *Critical thinking: Educational imperative*. (New Directions for Community Colleges, No. 77). San Francisco: Jossey-Bass.

Baxter Magolda, M. B. (1987). The affective dimension of learning: Faculty-student relationships that enhance intellectual development. *College Student Journal*, 21, 46-58.

Beck, S.E., Bennett, A., McLeod, R., and Molyneaux, D. (1992). Review of research on critical thinking in nursing education. *Review of Research in Nursing Education*, V, 1-30.

Benware, C.A. and Deci, E.L. (1984). Quality of learning with and active versus passive motivational set. *American Educational Research Journal*, 21 (4), 765-75.

Bowen, H.R. (1977). *Investment in learning: The individual and social value of American higher education*. San Francisco: Jossey-Bass.

Braxton, J., and Nordvall, R. (1985). Selective liberal arts colleges: Higher quality as well as higher prestige? *Journal of Higher Education*, 56, 538-554.

Campione, J.C., and Armbruster, B.B. (1985). Acquiring information from texts: an analysis of four approaches. In J. W. Segal, S. F. Chipman, and R. Glaser (Eds.), *Thinking and learning skills*, VI, 317-362. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cano, J. (1993). An assessment of the level of cognitive performance and critical thinking ability of selected agricultural education students. *Journal of Agricultural Education*, Summer, 25-31.

Cano, J., and Martinez, C. (1991). The relationship between cognitive performance and critical thinking abilities among selected agricultural education students. *Journal of Agricultural Education*, Spring, 24-29.

Capoor, M., and Gelfman, A. (1988). *Assessment of college outcomes with the value-added approach*. Paper presented at the meeting of the Association for Institutional Research, Phoenix, AZ.

Centra, J., and Rock, D. (1971). College environments and student academic achievement. *American Educational Research Journal*, 8, 623-634.

Chaffee, J. (1992). Teaching critical thinking across the curriculum. *Critical thinking: Educational imperative*. (New Directions for Community Colleges, No. 77, pp. 25-35). San Francisco: Jossey-Bass.

Cheyney, L.V. (1991). *National tests: What other countries expect their students to know*. Washington, D.C.: National Endowment for the Humanities.

Chickering, A.W. (1969). *Education and identity*. San Francisco: Jossey-Bass.

Chickering, A.W., and Gamson, Z.F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, 39(7), 3-7.

Chipman, S.F., Segal, J.W., and Glaser, R. (Eds.) (1985). *Thinking and learning skills (Volume 2: Research and Open Questions)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, P. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51, 281-309.

Cole, C.C., Jr. (1982). *Improving instruction: Issues and alternatives for higher education*. ASHE-ERIC Higher Education Reports #2. Washington, D.C.: ERIC Higher Education Clearinghouse.

Collins, J., Brown, J.S., and Newman, S.E. (1986). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L.B. Resnick (Ed.), *Cognition and instruction: Issues and agendas*. Hillsdale, NJ: Erlbaum and Associates.

Daly, J.A. (1992). *Assessing speaking and listening: Preliminary considerations for a national assessment.* Paper prepared for the Second NCES Study Design Conference on Collegiate Skills Assessment, Washington, D.C.

DeLisi, R., and Staudt, J. (1980). Individual differences in college students' performance on formal operations tasks. *Journal of Applied Developmental Psychology*, 16, 121-131.

Diaz, I.R. and English, R.A. (1987). Preferred learning styles of community college career and transfer students. *Community/Junior College Quarterly*, 11, 227-238.

Dollar, R. (1992). College influence on analytical thinking and communication skills of part-time versus full-time students. *College Student Journal*, 273-279.

Dressel, P., and Mayhew, L. (1954). *General education: Explorations in evaluation.* Westport, CT: Greenwood Press.

Dumont, R.G. and Troelstrup, R.L. (1980). Exploring relationships between objective and subjective measures of instructional Outcomes. *Research in Higher Education*, 12(1), 37-51.

Elshout, J.J. and Veenman, M.V.J. (1992). Relation between intellectual ability and working method of predictors of learning. *Journal of Educational Research*, 85(3), 134-143.

Endo, J., and Harpel, R. (1982). The effect of student-faculty interaction on students' educational outcomes. *Research in Higher Education*, 16, 115- 138.

Eniaiyeji, P.A. (1983). The comparative effects of teacher-demonstration and self-paced instruction on concept acquisition and problem-solving skills of college level chemistry students. *Journal of Research in Science Teaching*, 20 (8), 795-801.

Ennis, R.H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44-48.

Ewell, P.T. (1988). Outcomes, assessment, and academic improvement: In search of usable knowledge. In J. Smart (Ed.), *Higher education: Handbook of theory and research*, 4, 53-108.

Ewell, P.T. and Jones, D.P. (1991). Actions matter: The case for indirect measures in assessing higher education's progress on the national education goals. *Journal of General Education*, 42(2), 123-148.

Facione, P.A. (1990). *Executive Summary of Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. Millbrae, CA: California Academic Press.

Facione, P.A. (1992). *Empirical methods of theory and tool development for the assessment of college level critical thinking and problem-solving*. Paper prepared for the Second NCES Study Design Conference on Collegiate Skills Assessment, Washington, D.C.

Feldman, K.A. and Newcomb, T.M. (1969). *The impact of college on students*. San Francisco: Jossey-Bass.

Fischer, C., and Grant, G. (1983). Intellectual levels in college classrooms. In C. Ellner and C. Barnes (Eds.), *Studies of college teaching: Experimental results, theoretical interpretations and new perspectives*. Lexington, MA: D.C. Heath.

Forrest, A., and Steele, J. (1982). *Defining and measuring general education knowledge and skills*. Iowa City: American College Testing Program.

Friedlander, J.H. (1970). The importance of quality of effort in predicting college student attainment. Unpublished Ph.D. Dissertation, Los Angeles: The University of California, Los Angeles.

Frost, S.H. (1991). Fostering the critical thinking of college women through academic advising and faculty contact. *Journal of College Student Development*, 32, 359-366.

Fulkerson, F.E. and Martin, G. (1981). Effects of exam frequency on student performance, evaluations of instructors, and test anxiety. *Teaching Psychology*, 8, 90-93.

Gaff, J.G. (1973). Making a difference: The impacts of faculty. *Journal of Higher Education*, 44, 605-622.

Gaff, J.G. (1983). *General education today: A critical analysis of controversies, practices and reforms*. San Francisco: Jossey-Bass.

Gagne, R.M. (1977). *The conditions of learning*. (3rd Ed.). New York: Holt, Rinehart and Winston.

Gamson, Z.F. and Associates (1984). *Liberating education*. San Francisco: Jossey-Bass.

Gamson, Z.F. and Poulsen, S.J. (1989). Inventories of good practice: The next step for the seven principles for good practice in undergraduate education. *AAHE Bulletin*, 42, 7-8.

Glaser, E.M. (1985). Critical thinking: Educating for responsible citizenship in a democracy. *National Forum*, 65(1), 24-27.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39(2), 93-104.

Gray, P. (1993). Engaging students' intellects: The immersion approach to critical thinking in psychology instruction. *Teaching of Psychology*, 20 (2), 68-74.

Halpern, D.F. (1992). *A national assessment of critical thinking skills in adults: Taking steps toward the goal.* Paper prepared for the Second NCES Study Design Conference on Collegiate Skills Assessment, Washington, D.C.

Johnson, T. and Butts, G. (1982). The relationships among college science student achievement, engaged time, and personal characteristics. *Journal of Research in Science Teaching*, 20, 357-366.

Jones, D.P. and Folger, J. (1993). *Use of fiscal policy to achieve states' educational goals.* Denver, CO: Education Commission of the States.

Jones, E.A. (1993a). *Critical thinking literature review.* University Park, PA: National Center on Postsecondary Teaching, Learning and Assessment.

Jones, E.A. (1993b). *Communication competence assessment.* University Park, PA: National Center on Postsecondary Teaching, Learning and Assessment, The Pennsylvania State University.

Jones, E.A. and Ratcliff, J.L. (1991). Which general education curriculum is better: Core curriculum or the distribution requirement? *Journal of General Education*, 40, 69-101.

Kalous, T.D. (1990). Helping college students change their learning styles. *JCARL*, 1(1), 54-63.

Keimig, R.T. (1983). Raising academic standards: A guide to learning improvement. *ASHE-ERIC Higher Education Research Report #4.* Washington, DC: ERIC Clearinghouse on Higher Education.

Knefelkamp, L.L. (1974). Developmental instruction: Fostering intellectual and personal growth in college students. Unpublished Doctoral Dissertation, University of Minnesota, St. Paul.

Kurfiss, J.G. (1988). Critical thinking: Theory, research, practice and possibilities. *ASHE-ERIC Higher Education Report #2*. Washington, DC: ERIC Clearinghouse on Higher Education.

Lenning, O.T., Ed. (1976). *Improving educational outcomes*. (New Directions for Higher Education, No. 16). San Francisco: Jossey-Bass.

Lenning, O.T., Munday, L.A., Johnson, O.B., Vander Well, A.R., and Brue, E.J. (1974). *The many faces of college success and their non-intellective correlates: The published literature through the decade of the sixties*. ACT Educational Programs and Services Monograph #15. Iowa City: The American College Testing Program.

Light, R.J. (1990). *The Harvard assessment seminars: Explorations with students and faculty about teaching, learning, and student life*. First Report. Cambridge, MA: Harvard Graduate School of Education and Kennedy School of Government.

Light, R.J. (1992). *The Harvard assessment seminars: Explorations with students and faculty about teaching, learning, and student life*. Second Report. Cambridge, MA: Harvard Graduate School of Education and Kennedy School of Government.

McKeachie, W.J. (1980). Class size, large classes, and multiple sections. *Academe*, 66, 24-27.

McKeachie, W.J., Pintrich, P.R., Lin, Y.G., and Smith, D. (1986). *Teaching and learning in the college classroom: A review of the research literature*. Ann Arbor, MI: National Center for Research to Improve Teaching and Learning (NCRIPTAL), University of Michigan.

McLeish, J. (1968). *The lecture method*. Cambridge Monographs on Teaching Methods #1. Cambridge, England: Cambridge Institute of Education.

McMillan, J.H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education*, 26(1), 3-29.

McPeck, J. (1981). *Critical thinking and education*. New York: St. Martin's Press.

Mentkowski, M. and Rogers, G.P. (1988). *Establishing the validity of measures of college student outcomes*. Milwaukee, WI: Office of Research and Evaluation, Alverno College.

Mentkowski, M., and Strait, M. (1983). *A longitudinal study of student change in cognitive development, learning styles, and generic abilities in an outcome-centered liberal arts curriculum*. (Final report to the National Institute of Education, Research Report No. 6). Milwaukee, WI: Alverno College, Office of Research and Evaluation.

Mines, R.A., King, P.M., Hood, A.B., and Wood, P.K. (1990). Stages of intellectual development and associated critical thinking skills in college students. *Journal of College Student Development*, 31, 538-547.

National Center for Higher Education Management Systems (1983). *Student outcomes questionnaires: An implementation handbook, second edition*. Boulder, CO: National Center for Higher Education Management Systems.

National Institute of Education, Study Group on the Conditions of Excellence in American Higher Education (1984). *Involvement in learning: Realizing the potential of American higher education*. Washington, DC: U.S. Government Printing Office.

Newell, G. (1984). Learning from writing in two content areas: A case study/protocol analysis. *Research in the Teaching of English*, 18(3), 265- 87.

Nummedal, S.G. (1991). *Designing a process to assess higher-order thinking skills in college graduates: Issues of concern*. Paper prepared for the First NCES Study Design Conference, Washington, D.C.

Office of Educational Research and Improvement (1988). *Class size and public policy: Politics and panaceas*. Washington, D.C.: U.S. Department of Education.

Ory, J., and Braskamp, L. (1988). Involvement and growth of students in three academic programs. *Research in Higher Education*, 28, 116-129.

Pace, C.R. (1979). *Measuring outcomes of college: Fifty years of findings and recommendations for the future*. San Francisco: Jossey-Bass.

Pace, C.R. (1984). *Measuring the quality of student experiences: An account of the development and use of the College Student Experiences Questionnaire*. Los Angeles: Higher Education Research Institute, University of California, Los Angeles.

Pace, C.R. (1985). Perspectives and problems in student outcomes research. In P.T. Ewell (Ed.), *Assessing educational outcomes*. (New Directions for Institutional Research, No. 47, pp. 7-18). San Francisco: Jossey-Bass.

Pace, C.R. (1987). *CSEQ: Test manual and norms*. Los Angeles: Center for the Study of Evaluation, The University of California, Los Angeles.

Pace, C.R. (1990). *The undergraduates: A report of their activities and progress in college in the 1980's*. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles.

Pascarella, E.T. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J. Smart (Ed.), *Higher Education: Handbook of Theory and Research*, 1, 1-61.

Pascarella, E.T. (1989). The development of critical thinking: Does college make a difference? *Journal of College Student Development*, 30, 19-26.

Pascarella, E.T. (1991). The impact of college on students: The nature of the evidence. *The Review of Higher Education*, 14(4), 453-466.

Pascarella, E.T., and Terenzini, P.T. (1978). Student-faculty informal relationships and freshman year educational outcomes. *Journal of Educational Research*, 71, 183-189.

Pascarella, E.T. and Terenzini, P.T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass.

Paul, R.W. (1984). Critical thinking: Fundamental to education for a free society. *Educational Leadership*, 4-14.

Paul, R.W. (1985). The critical thinking movement: A historical perspective. *National Forum*, 65(1), 2-3.

Paul, R.W. (1992). Critical thinking: What, why and how. *In Critical thinking: Educational imperative*. (New Directions for Community Colleges, No. 77, pp. 3-24). San Francisco: Jossey-Bass.

Paul, R.W., Nosich, G.M., and Fisher, A.E. (1992). *Epistemic structures, not psychological structures, are at the core of the assessment of higher order thinking*. Paper prepared for the Second NCES Study Design Conference on Collegiate Skills Assessment, Washington, D.C.

Perfetto, G.A., Bransford, J.D., and Franks, J.J. (1983). Constraints on access in a problem-solving context. *Memory and Cognition*, 11(1), 24-31.

Perkins, D., Jay, E., and Tishman, S. (1993). *Assessing Critical Thinking: A Framework for Measuring Critical Thinking and Problem-Solving Skills at the College Level*. Paper prepared for the Second NCES Study Design Conference on Collegiate Skills Assessment, Washington, D.C. (revised on the basis of reviewer comments).

Perry, W. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart and Winston.

Peterson, M.W. (1987). *Academic management practices survey for the research program on the organizational context for teaching and learning*. Ann Arbor: National Center for Research to Improve Postsecondary Teaching and Learning, The University of Michigan.

Pike, G.R. (1989). Background, college experiences, and the ACT-COMP Exam: Using construct validity to evaluate assessment instruments. *Review of Higher Education*, 131, 91-117.

Pike, G.R. (1991). The effects of background, coursework, and involvement on students' grades and satisfaction. *Research in Higher Education*, 32(1), 15-30.

Pike, G.R., and Banta, T.W. (1989). *Using construct validity to evaluate assessment instruments: A comparison of the ACT-COMP exam and the ETS academic profile*. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Pike, G.R., and Phillippi, R. (1988). *Relationships between self-reported coursework and performance on the ACT-COMP exam: An analysis of the generalizability of the differential coursework methodology*. Paper presented at the meeting of the Association for the Study of Higher Education, St. Louis.

Pintrich, P.R. (1988). A process-oriented view of student motivation and cognition. In J. Stark and L. Mets (Eds.), *Improving teaching and learning through research*. (New Directions for Institutional Research, No. 57, pp. 65-79). San Francisco: Jossey-Bass:

Pintrich, P.R. and Johnson, G.R. (1990). Assessing and improving students' learning strategies. In M.D. Svinicki (Ed.), *The changing face of college teaching*. (New Directions for Teaching and Learning, No. 42, pp. 83-92). San Francisco: Jossey-Bass.

Pintrich, P.R., McKeachie, W.J., and Smith, D. (1989). *The motivated strategies for learning questionnaire*. Ann Arbor, MI: National Center for Research to Improve Teaching and Learning (NCRIPTAL), University of Michigan.

Porter, O.T. (1982). The role of quality of effort in defining institutional environments: An attempt to understand college uniqueness. Unpublished Ph.D. Dissertation. Los Angeles: The University of California, Los Angeles.

Postsecondary Education Planning Commission (1990). *The impact of undergraduate enhancement funding, report 7*. Tallahassee, Fla.: Postsecondary Education Planning Commission, State of Florida.

Ratcliff, J.L. (Ed.) (1993). *Assessment and curricular reform*. University Park, PA: National Center on Postsecondary Teaching, Learning, and Assessment, The Pennsylvania State University.

Ratcliff, J.L., and Associates. (1988). *Development and testing of a cluster-analytic model for identifying coursework patterns associated with general learned abilities of students (Progress Report No. 6)*. Ames: Iowa State University, College of Education.

Robertshaw, D. and Wolfle, L. (1983). Discontinuities in schooling and educational attainment. *Higher Education*, 12, 1-18.

Schalock, H.D. (1976). Structuring processes to improve student outcomes. In O.T. Lenning (Ed.), *Improving educational outcomes*. (New Directions for Higher Education, No. 16, pp. 25-54). San Francisco: Jossey-Bass.

Smith, D. (1977). College classroom interactions and critical thinking. *Journal of Educational Psychology*, 69, 180-190.

Smith, D. (1981). *Instruction and outcomes in an undergraduate setting*. Paper presented at the meeting of the American Educational Research Association, Los Angeles.

Smith, W.L., Hull, G.A., Land, R.E., Moore, M.T., Ball, C., Dunham, D.E., Hickey, L.S., and Ruzich, C.W. (1985). Some effects of varying the structure of a topic on college students' writing. *Written Communication*, 2(1), 73-89.

Steele, J. (1986). *Assessing reasoning and communication skills of postsecondary students*. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Stephenson, B.W. and Hunt, C. (1977). Intellectual and ethical development: A dualistic curriculum intervention for college students. *The Counseling Psychologist*, 6(4), 39-42.

Stone, N.R. (1990). Developing critical thinkers: Content and process. *Journal of Developmental Education*, 13(3), 20-22.

Terenzini, P.T. and Pascarella, E.T. (1991). Twenty years of research on college students: Lessons for future research. *Research in Higher Education*, 32(1), 83-92.

Terenzini, P.T., Pascarella, E.T., and Lorang, W. (1982). An assessment of the academic and social influences on freshman year educational outcomes. *Review of Higher Education*, 5, 86-110.

Terenzini, P.T., and Wright, T. (1987). Influences on students' academic growth during four years of college. *Research in Higher Education*, 26, 161-179.

Thomas, J.W., Rohwer, W.D., Jr. (1986). Academic studying: The role of learning strategies. *Educational Psychologist*, 21, 19-41.

Thorndike, R.M. (1990). The Washington State assessment experience. *Assessment Update*, 2(2), 7-9.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45, 89-125.

Vandament, W.E. (1991). *Registry of higher education reform*. Los Angeles: Council for the Development of Education.

Weinstein, M. (1990). Towards a research agenda for informal logic and critical thinking. *Informal Logic: Reasoning and Argumentation in Theory and Practice*, XII(3), 121-143.

Weinstein, C.E., Schulte, A.C., and Palmer, D.R. (1987). *Learning study strategies inventory (LASSI)*. Clearwater, FL: H&H Publishing.

Whitla, D.K. (1978). *Value added: Measuring the impact of undergraduate education*. Cambridge, MA: Office of Instructional Research and Evaluation, Harvard University.

Widick, C., Knefelkamp, L.L., and Parker, C.A. (1975). The counselor as a developmental instructor. *Counselor Education and Supervision*. (June), 286-296.

Wiggins, G.C. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.

Williams, D., Cook, P., Quinn, B., and Jensen, R. (1985). University class size: Is smaller better? *Research in Higher Education*, 23(3), 307-318.

Wilson, R., Wood, L., and Gaff, J. (1974). Social-psychological accessibility and faculty-student interaction beyond the classroom. *Sociology of Education*, 47, 74-92.

Winter, D.G., McClelland, D.C., and Stewart, A.J. (1981). *A new case for the liberal arts: Assessing institutional goals and student development*. San Francisco: Jossey-Bass.

Zemsky, R. (1989). *Structure and coherence: Measuring the undergraduate curriculum*. Philadelphia: Institute for Research on Higher Education, The University of Pennsylvania.

Zimmerman, A.P. (1991). A capstone problem solving/systems course at a two-year technical college. *NACTA Journal*, March, 26-29.

REFERENCES CONSULTED BUT NOT DIRECTLY CITED

Ayres, Q.W. and Bennett, R.W. (1983). University characteristics and student achievement. *Journal of Higher Education*, 54(5), 516-32.

Biggs, J. (1988). The role of metacognition in enhancing learning. *Australian Journal of Education*, 32(2), 127-138.

Bisanz, G.L., Vesonder, G.T., and Voss, J.F. (1978). Knowledge of one's own responding and the relation of such knowledge to learning. *Journal of Experimental Child Psychology*, 25, 116-128.

Border, L. and Van Note Chism, N. (Eds.). (1992). *Teaching for diversity*. (New Directions for Teaching and Learning, No. 49). San Francisco: Jossey-Bass.

Bowen, C.W. (1990). Representational systems used by graduate students while problem solving in organic synthesis. *Journal of Research in Science Teaching*, 27(4), 351-370.

Dumont, R.G. and Troelstrup, R.L. (1981). Measures and predictors of educational growth with four years of college. *Research in Higher Education*, 14 (1), 31-47.

Dunlop, D.L. and Fazio, F. (1976). Piagetian theory and abstract preferences of college science students. *JCST* (May), 297-300.

Eisert, D.C. and Tomlinson-Keasey, C. (1978). Cognitive and interpersonal growth during the college freshman year: A structural analysis. *Perceptual and Motor Skills*, 46, 995-1005.

Falcione, P.A. (1986). Testing college-level critical thinking. *Liberal Education*, 72 (3), 221-231.

Freie, J.F. Thinking and believing. *Point/Counterpoint*, 35 (3), 89-91.

Gaff, J.G. (1973). Making a difference: The impacts of faculty. *Journal of Higher Education*, 44 (8), 605-622.

Geiger, M.A., Boyle, E.J., and Pinto, J. (1992). A factor analysis of Kolb's revised learning style inventory. *Educational and Psychological Measurement*, 52, 753-759.

Heller, S. (1987). Ineffective teaching and fuzzy assignments can thwart "critical thinking" by college students, scholars argue. *Chronicle of Higher Education*, December 2, A13, A20.

Higbee, J.L., Ginter, E.J., and Taylor, W.D. (1991). Enhancing academic performance: Seven perceptual styles of learning. *RTDE*, 7, 5-10.

Hill, D.M. and Obenauf, P.A. (1979). Spatial visualization, problem solving, and cognitive development in freshman teacher education students. *Science Education*, 63 (5), 665-670.

Iverson, B.K., Pascarella, E.T., and Terenzini, P.T. (1984). Informal faculty-student contact and commuter college freshmen. *Research in Higher Education*, 21 (2), 123-136.

Keeley, S.M., Browne, M.N. and Kreutzer, J.S. (1982). A comparison of freshmen and seniors on general and specific essay tests of critical thinking. *Research in Higher Education*, 17 (2), 139-154.

King, P.M., Wood, P.K., and Mines, R.A. (1990). Critical thinking among college and graduate students. *The Review of Higher Education*, 13 (2), 167-186.

Kooker, E.W. (1971). The relationship between performance in a graduate course in statistics and the Miller analogies test and the Watson-Glaser critical thinking appraisal. *Journal of Psychology*, 77, 165-169.

Kuehner, A. (1992): Getting students to wage the battle in freshman composition. *TETYC* (February), 45-48.

Kuh, G.D., Bean, J.P., Bradley, R.K., Coomes, M.D., and Hunter, D.E. (1986). Changes in research on college students published in selected journals between 1969 and 1983. *The Review of Higher Education*, 9 (2), 177-192.

Kurland, D.J. (1991). The student's view of the text: Implication for reading and writing. *College Composition and Communication*, 42(4), 341-346.

McMurray, M.A. and Beisenherz, P. (1991). Reliability and concurrent validity of a measure of critical thinking skills in biology. *Journal of Research in Science Teaching*, 28(2), 183-192.

Miles, C. The fourth "r": Checklists for assessing thought in action. *Journal of Developmental Education*, 22, 32-33.

Miller, C.D., Alway, M., and McKinley, D.L. (1987). Effects of learning styles and strategies on academic success. *Journal of College Student Personnel*, 399-404.

Mintzes, J.J. (1982). Prior knowledge and locus of control in cognitive learning among college biology students. *Education*, 100(2), 138-145.

Novak, J.A. and Dettloff, J.M. (1989). Developing critical thinking skills in community college students. *JCST* (September/October), 22-25.

Ommen, J.L., Brainard, S.R., and Canfield, A.A. (1979). Learning preferences of younger and older students. *Community College Frontiers*, 29-33.

Pascarella, E.T. (1984). College environmental influences on students' educational aspirations. *The Journal of Higher Education*, 55 (6), 751-771.

Pascarella, E.T. (1985). Students' affective development within the college environment. *The Journal of Higher Education*, 56 (6), 640-663.

Paul, R. (1988). *Critical thinking and learning: A maximalist versus minimalist approach.* Paper presented at 21st Annual Conference of Western College Reading and Learning Association, Sacramento, California.

Peterson, M.W., Cameron, K.S., Mets, L.A., Jones, P., and Ettington, D. (1986). *The organizational context for teaching and learning: A review of the research literature.* Ann Arbor: National Center for Research to Improve Postsecondary Teaching and Learning, The University of Michigan.

Phye, G.D. and Sanders, C.E. (1992). Accessing strategic knowledge: individual differences in procedural and strategy transfer. *Contemporary Educational Psychology*, 17, 211-223.

Pike, G.R. (1992). The components of construct validity: A comparison of two measures of general education outcomes. *Journal of General Education*, 41, 130-159.

Ratcliff, R. and McKoon, G. (1989). Information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, 21, 139-155.

Roth, W. and Milkent, M.M. (1991). Factors in the development of proportional reasoning strategies by concrete operational college students. *Journal of Research in Science Teaching*, 28 (6), 553-566.

Simon, A. and Ward, L.O. (1974). The performance on the Watson-Glaser critical thinking appraisal of university students classified according to sex, type of course pursued, and personality score category. *Educational and Psychological Measurement*, 34, 957-960.

Slattery, P. (1990). Encouraging critical thinking: A strategy for commenting on college papers. *College Composition and Communications*, 41 (October), 332-335.

Stark, J.S. and Lowther, M.A. (1986). *Designing the learning plan: A review of research and theory related to college curricula.* Ann Arbor: National Center for Research to Improve Postsecondary Teaching and Learning, The University of Michigan.

Van Rossum, E.J. and Schenk, S.M. (1984). The relationship between learning conception, study strategy and learning outcome. *Journal of Educational Psychology*, 54, 73-83.

Warr, P.B., and Smith, J.S. (1970). Combining information about people: Comparisons between six models. *Journal of Personality and Social Psychology*, 16(1), 55-65.

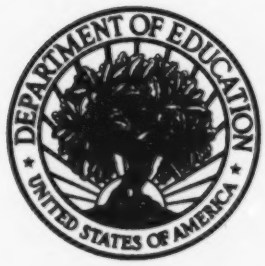
Wales, C.E. and Nardi, A.H. (1988). Do your students think or do they memorize? *Engineering Education*, 24, 682-688.

United States
Department of Education
Washington, DC 20208–5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Third Class



BEST COPY AVAILABLE